BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/29, 15/82, 5/10, A01H 5/00 — A2

(11) International Publication Number: WO 98/54328

(43) International Publication Date: 3 December 1998 (03.12.98)

(21) International Application Number: PCT/US98/11278

(22) International Filing Date: 2 June 1998 (02.06.98)

(30) Priority Data:
08/867,087 2 June 1997 (02.06.97) US

(71) Applicant: WASHINGTON STATE UNIVERSITY RESEARCH FOUNDATION [US/US]; N.E. 1615 Eastgate Boulevard, Pullman, WA 99164–1802 (US).

(72) Inventor: AN, Gynheung; Department of Life Science, Pohang University of Science and Technology, Pohang 790–784 (KR).

(74) Agent: STEPHENS, Donald, L., Jr.; Klarquist, Sparkman, Campbell, Leigh & Whinston, LLP, One World Trade Center, Suite 1600, 121 S.W. Salmon Street, Portland, OR 97204 (US).

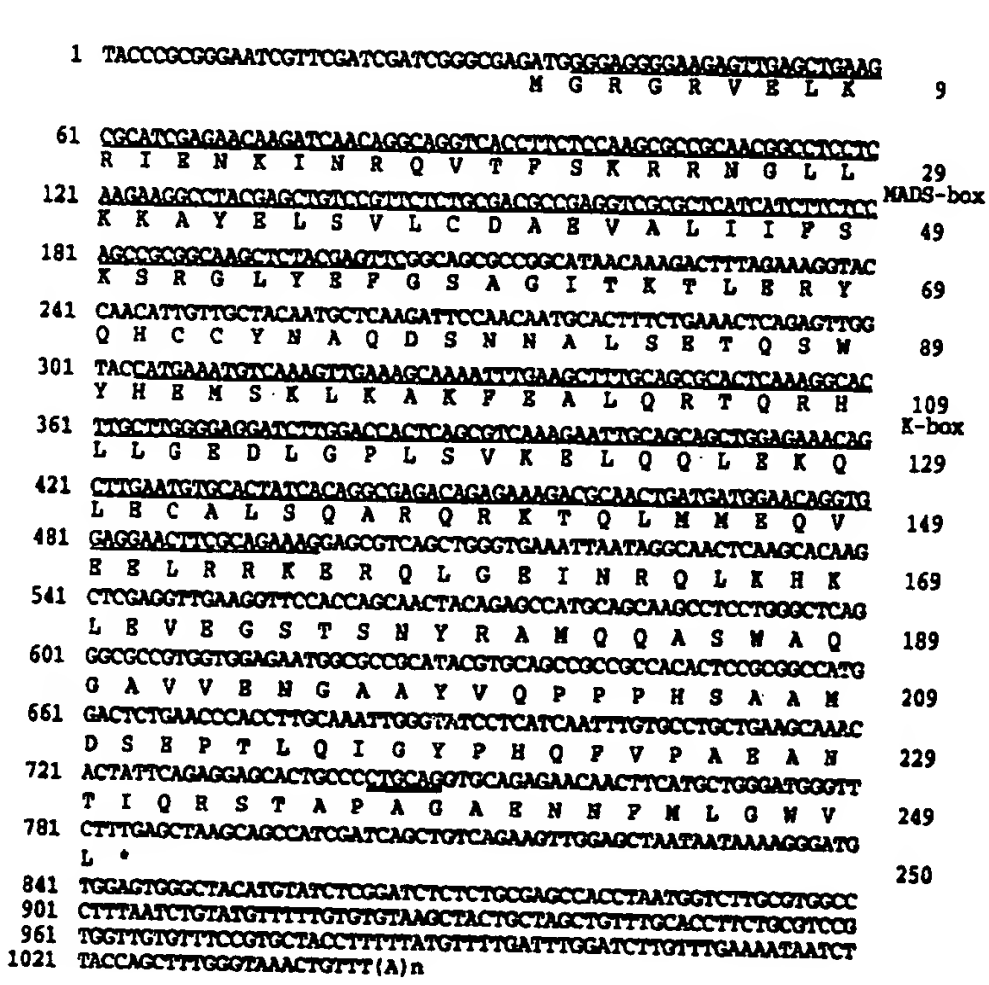
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ; TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: GENES CONTROLLING FLORAL DEVELOPMENT AND APICAL DOMINANCE IN PLANTS

(57) Abstract

The present invention provides compositions and methods for affecting the transition from vegetative to reproductive growth in a wide variety of plants. Several MADS–box genes, namely OsMADS6–8, have been isolated that, when expressed in transgenic plants, result in such phenotypes as, for example, reduced apical dominance or dwarfism and early flowering.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

GENES CONTROLLING FLORAL DEVELOPMENT AND APICAL DOMINANCE IN PLANTS

TECHNICAL FIELD

This invention is related to compositions and methods for affecting plant floral development and the timing of the transition from vegetative to reproductive growth.

BACKGROUND ART

Floral initiation is controlled by several factors including photoperiod, cold treatment, hormones, and nutrients (Coen, Plant Mol. Biol. 42:241-279, 1991; Gasser, Annu. Rev. Plant Physiol. Plant Mol. Biol. 42:621-649, 1991). Physiological studies have demonstrated that vegetative tissues are the site for signal perception and for generation of chemicals that cause the transition from vegetative growth to flowering (Lang, in: Encyclopedia of Plant Physiology, vol. 15, Berlin, ed., Springer-Verlag, pp. 1371-1536, 1965; Zeevaartm, in: Light and the Flowering Process, Vince-Prue et al., eds., Orlando Academic Press, pp. 137-142, 1984). Genetic analysis revealed that there are several types of mutants that alter flowering time. In *Arabidopsis thaliana*, there are at least two mutant groups based on their response to photoperiod and vernalization (Martinez-Zapater et al., in: *Arabidopsis*, Meyerowitz and Somerville, eds., Plainview, N.Y., Cold Spring Harbor Laboratory, pp. 403-433, 1994). These phenotypes suggest that there are multiple pathways that lead to flowering.

Study on mutants that interfere with normal flower development has provided some information on controlling the mechanisms of the development. This has led to the knowledge that there are at least two genes needed for induction of flower development: LEAFY (LFY) and APETALAI (API) genes in *Arabidopsis* (Weigel, Annu. Rev. Genet. 29:19-39, 1995), and FLORICAULA (FLO) and SQUAMOSA (SQUA) genes in *Antirrhinum majus* (Bradley et al., Cell 72:85-95, 1993). Cloning and analysis of these genes revealed that the LFY and FLO genes are homologs and encode proteins that each contain a proline-rich region at the N-terminus and a highly acidic central region, which are features of certain types of transcription factors that contain a conserved MADS-box sequence (Huijser et al., EMBO J. 11:1239-1249, 1992; Mandel et al., Nature 360:273-277, 1992). MADS box-containing genes were isolated from several plant species and are known to play important roles in plant development, especially flower development. *Arabidopsis* homeotic genes - AGAMOUS (AG), PISTILATA (PI), and APETALA3 (AP3) - are members of the MADS box gene family (Yanofsky et al., Nature 346:35-39, 1990; Goto and Meyerowitz, Genes Devel. 8:1548-1560, 1994; Jack et al., Cell 68:683-697, 1992). Similar homeotic genes from *A. majus*, namely PLENA (PLE), GLOBOSA (GLO), and DEFICIENS A (DEFA), are also MADS box genes (Bradley et al., Cell 72:85-95, 1993; Tröbner et al., EMBO J. 11:4693-4704, 1992; Sommer et al., EMBO J. 9:605-613, 1990). Characterization of these gene

products showed that the conserved MADS box domain is for sequence-specific DNA binding, dimerization, and attraction of secondary factors (Pellegrini et al., Nature 376:490-498, 1995). The DNA sequence with which the MADS box domains interact is the consensus finding site, CCA/T$_6$GG (Pollock and Treisman, Genes Dev. 5:2327-2341, 1991; Huang et al., Nucl. Acids Res. 21:4769-4776, 1993). In addition to the MADS-box domain, the plant MADS box proteins include the K-box domain, a second conserved region carrying 65-70 amino acid residues. The K-box domain was named due to the structural resemblance to the coiled coil domain of keratin (Ma et al., Genes Dev. 5:484-495, 1991) and has been suggested to be related to protein-protein interactions (Pnueli et al., Plant J. 1:255-266, 1991). Similar MADS-box genes have also been studied in other plants including tomato, rape, tobacco, petunia, maize, and rice (Theißen and Saedler, Curr. Opin. Genet. Dev. 5:628-639, 1995). A number of plant MADS box genes that deviate from the functions of the typical meristem identity and organ identity genes have been identified. These genes are involved in the control of ovule development (Angenent et al., Plant Cell 7:1569-1582, 1995), vegetative growth (Mandel et al., Plant Mol. Biol. 25:319-321, 1994), root development (Rounseley et al., Plant Cell 7:1259-1269, 1995), embryogenesis (Heck et al., Plant Cell 7:1271-1282, 1995), or symbiotic induction (Heard and Dunn, Proc. Natl. Acad. Sci. USA 5273-5277, 1995).

There are a large number of MADS box genes in each plant species. In maize, a multigene family consists of at least 50 different MADS box genes and these genes are dispersed throughout the plant genome (Mena et al., Plant J. 8:845-854, 1995; Fischer et al., Proc. Natl. Acad. Sci. USA 92:5331-5335, 1995). The MADS box multigene family can be divided into several subfamilies according to their primary sequences, expression patterns, and functions (Theißen and Saedler, Curr. Opin. Genet. Dev. 5:628-639, 1995).

Several MADS genes from *Oryza sativa* have been identified and sequenced, including *OsMADS1* (Chung et al., Plant Mol. Bio. 26:657-665, 1994; Chung et al., Plant Sci. 109:45-56 1995; WO 96/11566) and *OsMADS5* (Kang et al., Plant Mol. Biol. 29:1-10, 1997). The present invention is directed to several other distinct MADS genes of *Oryza sativa*.

The timing of the transition from vegetative growth to flowering is one of the most important steps in plant development. This step determines the quality and quantity of most crop species by affecting the balance between vegetative and reproductive growth. It would therefore be highly desirable to have means to affect the timing of this transition. The present invention meets these needs and others.

SUMMARY OF THE INVENTION

The present invention provides, *intra alia*, compositions and methods related to the *OsMADS6*, *OsMADS7*, and *OsMADS8* genes of *Oryza sativa* and alleles and homologs of such

genes. Expression of such genes in transgenic plants causes an altered phenotype, including phenotypes related to the timing of the transition between vegetative and reproductive growth.

It is an object of the invention to provide isolated nucleic acids representing MADS genes and alleles that, when expressed in transgenic plants, confer on the plants at least one phenotype including (1) diminished apical dominance, (2) early flowering, (3) a partially or completely altered daylength requirement for flowering, (4) greater synchronization of flowering, or (5) a relaxed vernalization requirement.

It is another object of the invention to provide isolated nucleic acids comprising (1) a sequence of at least 30 contiguous nucleotides of *OsMADS6* (SEQ ID NO:2) or an allele or homolog thereof, or (2) a sequence of at least 100 contiguous nucleotides that has at least 70% nucleotide sequence similarity with *OsMADS6*. When expressed in a transgenic plant, such nucleic acids produce at least one phenotype including (1) diminished apical dominance, (2) early flowering, (3) a partially or completely altered daylength requirement for flowering, (4) greater synchronization of flowering, or (5) a relaxed vernalization requirement. Preferably, such isolated nucleic acids comprise only silent or conservative substitutions to a native (wild-type) gene sequence.

A further object of the invention is to provide transgenic plants comprising such nucleic acids.

A further object of the invention is to provide probes and primers comprising a fragment of the *OsMADS6* gene, the probes and primers being capable of specifically hybridizing under stringent conditions to the *OsMADS6* gene. Such probes and primers are useful, for example, for obtaining homologs of the *OsMADS6* gene from plants other than rice.

It is a further object of the invention to use the nucleic acids described above to produce transgenic plants having altered phenotypes, specifically, to introduce such nucleic acids into plant cells, thereby producing a transformed plant cell, and to regenerate from the transformed plant cell a transgenic plant comprising the nucleic acid.

The foregoing and other objects and advantages of the invention will become more apparent from the following detailed description and accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows nucleotide and deduced amino-acid sequences of the *OsMADS6* cDNA (SEQ ID NO:2 and SEQ ID NO:3). MADS-box and K-box regions are underlined. The positions of nucleotides and amino acids are shown on the left and right, respectively. The double underlined sequence is the *Pst*I site, which was used to generate the gene-specific probe of the 360 bp fragment located at the 3' region of the *OsMADS6* cDNA.

FIG. 2 shows nucleotide and deduced amino-acid sequences of the *OsMADS7* cDNA (SEQ ID NO:4 and SEQ ID NO:5). The MADS-box and K-box regions are underlined. The positions of

nucleotides and amino acids are shown on the left and right, respectively. The double-underlined sequence is the *Pst*I site, which was used to generate the gene-specific probe of the 280 bp fragment located at the 3'-end region of the *OsMADS7* cDNA.

FIG. 3 shows nucleotide and deduced amino-acid sequences of the *OsMADS8* cDNA (SEQ ID NO:6 and SEQ ID NO:7). The MADS-box and K-box regions are underlined. The positions of nucleotides and amino acids are shown on the left and right, respectively. The double-underlined sequence is the *Nhe*I site, which was used to generate the gene-specific probe of the 230-bp fragment located at 3'-end region of the *OsMADS8* cDNA.

FIGS. 4A-4C show alignments of MADS-box, K-box, and C-terminal regions of OsMADS6, OsMADS7, and OsMADS8 proteins (SEQ ID NO:3, 5, and 7, respectively) with other MADS-box proteins. Gaps were introduced for optimal alignment. FIG. 4A shows alignments of MADS-box regions. FIG. 4B shows alignment of K-box regions. FIG. 4C shows alignment of C-terminal regions. **1**, OsMADS6 (rice) (SEQ ID NO:3: MADS box, K box, and C-terminal end]; **2**, ZAG3 (maize) (SEQ ID NO:8, MADS box; SEQ ID NO:20, K box; SEQ ID NO:32, C-terminal end); **3**, ZAG5 (maize) (SEQ ID NO:9, MADS box; SEQ ID NO:21, K box; SEQ ID NO:33, C-terminal end); **4**, AGL6 (*Arabidopsis*) (SEQ ID NO:10, MADS box; SEQ ID NO:22, K box; SEQ ID NO:34, C-terminal end); **5**, OsMADS8 (rice) (SEQ ID NO:7, MADS box, K box, and C-terminal end); **6**, OsMADS7 (rice) (SEQ ID NO:5, MADS box, K box, and C-terminal end); **7**, FBP2 (petunia) (SEQ ID NO:11, MADS box; SEQ ID NO:23, K box; SEQ ID NO:35, C-terminal end); **8**, TM5 (tomato) (SEQ ID NO:12, MADS box; SEQ ID NO:24, K box; SEQ ID NO:36, C-terminal end); **9**, OM1 (orchid) (SEQ ID NO:13, MADS box; SEQ ID NO:25, K box; SEQ ID NO:37, C-terminal end); **10**, AGL2 (*Arabidopsis*) (SEQ ID NO:14, MADS box; SEQ ID NO:26, K box; SEQ ID NO:38, C-terminal end); **11**, AGL4 (*Arabidopsis*) (SEQ ID NO:15, MADS box; SEQ ID NO:27, K box; SEQ ID NO:39, C-terminal end); **12**, OsMADS1 (rice) (SEQ ID NO:1, MADS box, K box, and C-terminal end); **13**, AP1 (*Arabidopsis*) (SEQ ID NO:16, MADS box; SEQ ID NO:28, K box; SEQ ID NO:40, C-terminal end); **14**, AG (*Arabidopsis*) (SEQ ID NO:17, MADS box; SEQ ID NO:29, K box; SEQ ID NO:41, C-terminal end); **15**, AP3 (*Arabidopsis*) (SEQ ID NO:18, MADS box; SEQ ID NO:30, K box; SEQ ID NO:42, C-terminal end); **16**, PI (*Arabidopsis*) (SEQ ID NO:19, MADS box; SEQ ID NO:31, K box; SEQ ID NO:43, C-terminal end).

FIG. 5 shows genetic maps of the *OsMADS* genes. The locations of *OsMADS* genes along with RFLP markers (RG, G), cDNA markers (RZ and C), and microsatellite markers (RM) are indicated. Map distance is given in cM on the left of each chromosome. Dark bars represent the centromere regions.

DETAILED DESCRIPTION OF THE INVENTION

Three different MADS-box genes have been isolated from rice, *OsMADS6*, *OsMADS7*, and *OsMADS8*. *OsMADS6-8* were isolated from cDNA libraries under moderate stringency hybridization conditions using *OsMADS1* as a probe, as described below. Details regarding the isolation, characterization, and nucleotide sequence are provided in (Chung et al., Plant Mol. Bio. 26:657-665, 1994; Chung et al., Plant Sci. 109:45-56, 1997; WO 96/11566) and *OsMADS5* (Kang et al., Plant Mol. Biol. 29:1-10, 1997).

The present invention provides compositions and methods related to the MADS-box genes of rice, *OsMADS6*, *OsMADS7*, *OsMADS8*, and their alleles and homologs (collectively referred to below as "MADS genes"). These genes are useful, for example, for producing dwarf plants and for affecting the timing of the transition from vegetative to reproductive growth in a wide variety of plants, including various dicotyledonous and monocotyledonous crop plants and tree species (see Schwarz-Sommer et al., Science 250:931-936, 1990, regarding "MADS-box" genes).

Use of the Genes and their Alleles and Homologs for Crop Improvement

The MADS genes and polypeptides disclosed herein are useful for the following purposes, among others.

Early flowering. The timing of the transition between vegetative and reproductive growth is an important agronomic trait, serving as a crucial factor in determining crop yields. Expression of MADS genes in transgenic plants makes it possible to affect the transition from vegetative to reproductive growth in a wide variety of plants, whether the plants are long-day, short-day, or day-neutral plants.

When MADS genes are expressed in transgenic plants of day-neutral species, the resulting transgenic plants flower earlier than control plants. Transgenic long-day and short-day flowering plants expressing the MADS genes also flower earlier under permissive conditions than control plants. The compositions and methods according to the present invention therefore permit one to reduce the length of the vegetative growth stage of cereal, fruit, vegetable, floricultural, and other crop species.

Producing dwarf plant varieties. Although it has been possible to enhance the harvest index in grain crops by the use of dwarfing genes, the isolation of these genes producing dwarf phenotypes has been difficult.

Transgenic plants comprising a MADS transgene are shorter than controls. Therefore, MADS genes are useful for producing dwarf plant varieties for a variety of plants including cereal, fruit, and floricultural species.

Synchronizing reproductive growth. Transgenic plants expressing an transgene flower more synchronously than controls. Therefore, the gene can be used for crops for which synchronized harvesting is economically beneficial, allowing more effective use of mechanized

harvesting of fruit species or the production of floricultural species having improved flower quality, for example.

Producing day-neutral plant varieties. Expression of an transgene in daylength-sensitive (i.e., long-day or short-day) plants at least partially overrides the photoperiod requirement for flowering and can completely override the photoperiod requirement. By introducing such a transgene into a wide variety of photoperiod-sensitive crop or floricultural species, including, but not limited to rice, soybeans, chrysanthemums and orchids, these plants effectively become day-neutral, permitting multiple crops to be grown per year. For example, flowers can be induced year-round by introducing the transgene into floricultural species such as chrysanthemum and orchid.

Delaying flowering and fruiting. By suppressing the expression of a native MADS gene by conventional approaches (e.g., by employing antisense, co-suppression, gene replacement, or other conventional approaches to suppressing plant gene expression), flowering and fruiting can be delayed. Delayed reproductive growth can thereby increase the length of the vegetative growth stage and cause the plants to grow faster since the energy used for development of flowers and seeds can be saved for vegetative growth. Thus, delaying or eliminating reproductive growth results in a higher yield of vegetable species such as spinach, radish, cabbage, or tree species. In addition, such plants will be more desirable as garden and street species, since their production of pollen allergens can be reduced or eliminated.

Overcoming the vernalization requirement. *MADS* genes are useful for overriding the vernalization requirement of certain plant species. Induction of flowering of transgenic plants that constitutively express a *MADS* gene thus becomes insensitive to temperature.

Growing plants in space. Plants grown extraterrestrially are preferably insensitive to photoperiod and temperature for flowering. Transgenic plants carrying *MADS* genes would be expected to flower in the extremely abnormal growth conditions found in a space shuttle or space station.

Cloning and analysis of alleles and homologs. The availability of MADS genes makes it possible to obtain alleles and homologs of these genes by conventional methods, through the use of nucleic acid and antibody probes and primers, as discussed below.

DEFINITIONS AND METHODS

The following definitions and methods are provided to better define the present invention and to guide those of ordinary skill in the art in the practice of the present invention. Definitions of common terms in molecular biology may also be found in Rieger et al., Glossary of Genetics: Classical and Molecular, 5th edition, Springer-Verlag: New York, 1991; and Lewin, Genes VI, Oxford University Press: New York, 1997.

The term "plant" encompasses any plant and progeny thereof. The term also encompasses parts of plants, including seed, cuttings, tubers, fruit, flowers, etc.

A "reproductive unit" of a plant is any totipotent part or tissue of the plant from which one can obtain a progeny of the plant, including, for example, seeds, cuttings, buds, bulbs, somatic embryos, etc.

"Natural photoperiod conditions" are photoperiod (i.e., daylength) conditions as provided by sunlight at a given location, whether under field conditions. A photoperiod provided by artificial lighting but having a daylength approximating that of sunlight would also be considered a natural photoperiod condition.

Nucleic Acids

Nucleic acids useful in the practice of the present invention comprise at least one of the isolated genes disclosed herein, namely *OsMADS6*, *OsMADS7*, and *OsMADS8*, and their alleles, homologs, fragments, and variant forms thereof.

The term "MADS gene" for example, refers to a plant gene that contains a MADS-box sequence, and preferably also a K-box sequence, and that is associated with one or more of the following phenotypes when expressed as a transgene in transgenic plants: (1) diminished apical dominance (as shown, for example, by dwarf stature) and (2) early flowering, and can also be associated with, for example, (3) altered daylength requirement for flowering; (4) greater synchronization of flowering; and (5) relaxed vernalization requirement. The MADS gene encompasses the respective coding sequences and genomic sequences flanking the coding sequence that are operably linked to the coding sequence, including regulatory elements and/or intron sequences. Also encompassed are alleles and homologs.

The term "native" refers to a naturally-occurring nucleic acid or polypeptide, including a wild-type sequence and an allele thereof.

A "homolog" of a MADS gene is a native gene sequence isolated from a plant species other than the species from which the MADS gene was originally isolated and having similar biologically activities, e.g., dwarfism and early flowering.

"Isolated". An "isolated" nucleic acid has been substantially separated or purified away from other nucleic acid sequences in the cell of the organism in which the nucleic acid naturally occurs, i.e., other chromosomal and extrachromosomal DNA and RNA. The term "isolated" thus encompasses nucleic acids purified by standard nucleic acid-purification methods. The term also embraces nucleic acids prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids.

DNA constructs incorporating a MADS gene or fragment thereof preferably place the protein-coding sequence under the control of an operably linked promoter that is capable of expression in a plant cell. Various promoters suitable for expression of heterologous genes in plant

cells are known in the art, including constitutive promoters, e.g. the cauliflower mosaic virus (CaMV) 35S promoter, which is expressed in many plant tissues, organ- or tissue-specific promoters, and promoters that are inducible by chemicals such as methyl jasminate, salicylic acid, or Safener, for example.

Plant transformation and regeneration. In addition to the methods for plant transformation and regeneration described in the Examples below for making transgenic plants, other well-known methods can be employed.

Fragments, probes, and primers. A fragment of an *OsMADS* nucleic acid according to the present invention is a portion of the nucleic acid that is less than full-length and comprises at least a minimum length capable of hybridizing specifically with the corresponding *OsMADS* nucleic acid (or a sequence complementary thereto) under stringent conditions as defined below. The length of such a fragment is preferably at least 15 nucleotides in length, more preferably at least 30 nucleotides, more preferably at least 50 nucleotides, and most preferably at least 100 nucleotides.

A "probe" comprises an isolated nucleic acid attached to a detectable label or reporter molecule well known in the art. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

"Primers" are short nucleic acids, preferably DNA oligonucleotides 15 nucleotides or more in length, that can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, then extended along the target DNA strand by a polymerase, preferably a DNA polymerase. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods well known in the art. PCR-primer pairs can be derived from the sequence of a nucleic acid according to the present invention, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA).

Methods for preparing and using probes and primers are described, for example, in Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, ed. Sambrook et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; Current Protocols in Molecular Biology, ed. Ausubel et al., Greene Publishing and Wiley-Interscience, New York, 1987 (with periodic updates); and Innis et al., PCR Protocols: A Guide to Methods and Applications, Academic Press: San Diego, 1990.

Substantial similarity. A first nucleic acid is "substantially similar" to a second nucleic acid if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 70%-90% of the nucleotide bases, and preferably greater than 90% of the nucleotide bases. ("Substantial sequence complementarity" requires a similar degree of sequence complementarity.) Sequence similarity can be determined by comparing the nucleotide sequences of two nucleic acids

using sequence analysis software such as the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, Madison, WI.

Alternatively, two nucleic acids are substantially similar if they hybridize under stringent conditions, as defined below.

Operably linked. A first nucleic-acid sequence is "operably" linked with a second nucleic-acid sequence when the first nucleic-acid sequence is placed in a functional relationship with the second nucleic-acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein coding regions, in reading frame.

"Recombinant". A "recombinant" nucleic acid is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques.

Techniques for nucleic-acid manipulation are described generally in, for example, Sambrook et al. (1989) and Ausubel et al. (1987, with periodic updates).

Preparation of recombinant or chemically synthesized nucleic acids; vectors, transformation, host cells. Large amounts of a nucleic acid according to the present invention can be produced by recombinant means well known in the art or by chemical synthesis.

Natural or synthetic nucleic acids according to the present invention can be incorporated into recombinant nucleic-acid constructs, typically DNA constructs, capable of introduction into and replication in a host cell. Usually the DNA constructs will be suitable for replication in a unicellular host, such as *E. coli* or other commonly used bacteria, but can also be introduced into yeast, mammalian, plant or other eukaryotic cells.

Preferably, such a nucleic-acid construct is a vector comprising a replication system recognized by the host. For the practice of the present invention, well-known compositions and techniques for preparing and using vectors, host cells, introduction of vectors into host cells, etc. are employed, as discussed, *inter alia*, in Sambrook et al., 1989, or Ausubel et al., 1987.

A cell, tissue, organ, or organism into which has been introduced a nucleic acid according to an embodiment of the present invention, such as a recombinant vector, is considered "transformed" or "transgenic." A recombinant DNA construct that is present in a transgenic host cell, particularly a transgenic plant, is referred to as a "transgene." The term "transgenic" or "transformed" when referring to a cell or organism, also includes (1) progeny of the cell or organism and (2) plants produced from a breeding program employing such a "transgenic" plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of the recombinant DNA construct.

Conventional methods for chemical synthesis of nucleic acids are used, for example, in Beaucage and Carruthers, Tetra. Letts. 22:1859-1862, 1981, and Matteucci et al., J. Am. Chem. Soc. 103:3185, 1981. Chemical synthesis of nucleic acids can be performed, for example, on commercial automated oligonucleotide synthesizers.

Nucleic-Acid Hybridization: "Stringent Conditions"; "Specific". The term "stringent conditions" is functionally defined with regard to the hybridization of a nucleic-acid probe to a target nucleic acid (i.e., to a particular nucleic-acid sequence of interest) by the hybridization procedure discussed in Sambrook et al., 1989 at 9.52-9.55. *See also*, Sambrook et al., 1989 at 9.47-9.52, 9.56-9.58; Kanehisa, Nuc. Acids Res. 12:203-213, 1984; and Wetmur and Davidson, J. Mol. Biol. 31:349-370, 1968. According to one embodiment of the invention, "moderate stringency" hybridization conditions are hybridization at 60°C in a hybridization solution including 6x SSC, 5X Denhardt's reagent, 0.5% SDS, 100 μg/mL denatured, fragmented salmon sperm DNA, and the labeled probe (Sambrook et al., 1989), and "high stringency" conditions are hybridization at 65-68°C in the same hybridization solution.

Regarding the amplification of a target nucleic- acid sequence (e.g., by PCR) using a particular amplification primer pair, stringent conditions are conditions that permit the primer pair to hybridize only to the target nucleic-acid sequence to which a primer having the corresponding wild-type sequence (or its complement) would bind.

Nucleic-acid hybridization is affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of mismatched bases between the hybridizing nucleic acids.

When referring to a probe or primer, the term "specific for (a target sequence)" indicates that the probe or primer hybridizes only to the target sequence in a given sample comprising the target sequence.

Nucleic-acid amplification. As used herein, "amplified DNA" refers to the product of nucleic-acid amplification of a target nucleic-acid sequence. Nucleic-acid amplification can be accomplished by any of the various nucleic-acid amplification methods known in the art, including the polymerase chain reaction (PCR). A variety of amplification methods are known in the art and are described, *inter alia*, in U.S. Patent Nos. 4,683,195 and 4,683,202 and in PCR Protocols: A Guide to Methods and Applications, Innis et al. eds., Academic Press, San Diego, 1990.

In situ hybridization. A number of techniques have been developed in which nucleic-acid probes are used to locate specific DNA sequences on intact chromosomes *in situ*, a procedure called "*in situ* hybridization." See, e.g., Pinkel et al., Proc. Natl. Acad. Sci. USA 85:9138-9142, 1988 (regarding fluorescence *in situ* hybridization), and Lengauer et al., Hum. Mol. Genet. 2:505-512, 1993 (regarding "chromosomal bar codes"). Well-known methods for *in situ* hybridization and for the preparation of probes or primers for such methods are employed in the practice of the present invention, including direct and indirect *in situ* hybridization methods.

Methods of obtaining genomic clones, alleles, and homologs. Based upon the availability of the nucleotide sequences of the *MADS* genes disclosed herein, other MADS genes (e.g., alleles and homologs) and genomic clones corresponding thereto can be readily obtained from a wide variety of plants by cloning methods known in the art.

For example, one or more primer pairs can be used to amplify such alleles or homologs by the polymerase chain reaction (PCR). Alternatively, the disclosed cDNA of *OsMADS6*, *OsMADS7*, or *OsMADS8*, or a fragment thereof, can be used to probe a cDNA or genomic library made from a given plant species.

Nucleotide-Sequence and Amino-Acid Sequence Variants. "Variant" DNA molecules are DNA molecules containing minor changes to a native, or wild-type, sequence, i.e., changes in which one or more nucleotides of a native sequence are deleted, added, and/or substituted while substantially maintaining wild-type biological activity. Variant DNA molecules can be produced, for example, by standard DNA mutagenesis techniques or by chemically synthesizing the variant DNA molecule. Such variants do not change the reading frame of the protein-coding region of the nucleic acid.

Amino-acid substitutions are preferably substitutions of single amino-acid residues. DNA insertions are preferably of about 1 to 10 contiguous nucleotides and deletions are preferably of about 1 to 30 contiguous nucleotides. Insertions and deletions are preferably insertions or deletions from an end of the protein-coding or non-coding sequence (i.e., a truncation of the native sequence) and are preferably made in adjacent base pairs. Substitutions, deletions, insertions or any combination thereof can be combined to arrive at a final construct. For the sequences disclosed herein, amino acid substitutions preferably are located outside sequences that are conserved among *OsMADS6-8* and homologs thereof.

Preferably, variant nucleic acids according to the present invention are "silent" or "conservative" variants. "Silent" variants are variants of a native sequence in which there has been a substitution of one or more base pairs but no change in the amino-acid sequence of the polypeptide encoded by the sequence. "Conservative" variants are variants of a native sequence in which at least one codon in the protein-coding region of the native sequence has been changed, resulting in a conservative change in one or more amino-acid residues of the polypeptide encoded by the nucleic-acid sequence, i.e., an amino-acid substitution. A number of conservative amino-acid substitutions are listed in Table 1. In addition, there can be a substitution (resulting in a net gain or loss) of one or more cysteine residues, thereby affecting disulfide linkages in the encoded polypeptide.

TABLE 1

Original Residue	Conservative Substitutions
Ala	ser
Arg	lys
Asn	gln, his
Asp	glu
Cys	ser
Gln	asn
Glu	asp
Gly	pro
His	asn; gln
Ile	leu, val
Leu	ile; val
Lys	arg; gln; glu
Met	leu; ile
Phe	met; leu; tyr
Ser	thr
Thr	ser
Trp	tyr
Tyr	trp; phe
Val	ile; leu

Substantial changes in function are made by selecting substitutions that are less conservative than those in Table 1, i.e., selecting residues that differ more significantly in their effect on maintaining: (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation; (b) the charge or hydrophobicity of the molecule at the target site; or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in protein properties are those in which: (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histadyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

Polypeptides

The term "OsMADS protein" (or "OsMADS polypeptide") refers to a protein encoded by an OsMADS gene that has at least about 70% homology with a given native OsMADS polypeptide and preferably retains biological activity of the wild-type polypeptide. An OsMADS polypeptide can be isolated from a natural source, produced by the expression of a recombinant *OsMADS* nucleic acid, or be chemically synthesized by conventional methods, for example.

Polypeptide sequence homology. Ordinarily, the polypeptides encompassed by the present invention are at least about 90% homologous to a native polypeptide, and more preferably at least

about 95% homologous. Preferably, such polypeptides also have characteristic structural features and biological activity of the native polypeptide.

Polypeptide homology is typically analyzed using sequence analysis software such as the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, Madison, WI. Polypeptide sequence analysis software matches homologous sequences using measures of homology assigned to various substitutions, deletions, substitutions, and other modifications.

"Isolated," "Purified," "Homogeneous" Polypeptides. A polypeptide is "isolated" if it has been separated from the cellular components (nucleic acids, lipids, carbohydrates, and other polypeptides) that naturally accompany it. Such a polypeptide can also be referred to as "pure" or "homogeneous" or "substantially" pure or homogeneous. Thus, a polypeptide which is chemically synthesized or recombinant (i.e., the product of the expression of a recombinant nucleic acid, even if expressed in a homologous cell type) is considered to be isolated. A monomeric polypeptide is isolated when at least 60-90% by weight of a sample is composed of the polypeptide, preferably 95% or more, and more preferably more than 99%. Protein purity or homogeneity is indicated, for example, by polyacrylamide gel electrophoresis of a protein sample, followed by visualization of a single polypeptide band upon staining the polyacrylamide gel; high pressure liquid chromatography; or other methods known in the art.

Protein purification. The polypeptides of the present invention can be purified by any of the means known in the art. Various methods of protein purification are described, e.g., in Guide to Protein Purification, ed. Deutscher, Meth. Enzymol. 185, Academic Press, San Diego, 1990; and Scopes, Protein Purification: Principles and Practice, Springer Verlag, New York, 1982.

Variant forms of polypeptides; labeling. Variant polypeptides are those in which there have been substitutions, deletions, insertions or other modifications of a native polypeptide sequence. Variant polypeptides substantially retain structural and/or biological characteristics and are preferably silent or conservative substitutions of one or a small number of amino acid residues.

Wild-type polypeptide sequences can be modified by conventional methods, e.g., by acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, and labeling, whether accomplished by *in vivo* or *in vitro* enzymatic treatment of a native polypeptide or by protein synthesis using modified amino acids.

Any of a variety of conventional methods and reagents for labeling polypeptides and fragments thereof can be employed in the practice of the invention. Typical labels include radioactive isotopes, ligands or ligand receptors, fluorophores, chemiluminescent agents, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, e.g., in Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, ed. Sambrook et al., Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1989; and

Current Protocols in Molecular Biology, ed. Ausubel et al., Greene Publishing and Wiley-Interscience: New York, 1987 (with periodic updates).

Polypeptide Fragments. The present invention also encompasses polypeptide fragments that lack at least one residue of a native full-length polypeptide yet retain at least one of the biological activities characteristic of the native polypeptide. For example, the fragment can cause early flowering or dwarf phenotypes when expressed as a transgene in a host plant. An immunologically active fragment of a given full-length polypeptide is capable of raising antibodies specific for the full-length polypeptide in a target immune system (e.g., murine or rabbit) or of competing with the full-length polypeptide for binding to such specific antibodies, and is thus useful in immunoassays for the presence of the native polypeptide in a biological sample. Such immunologically active fragments typically have a minimum size of 7 to 17 amino acids.

Fusion polypeptides. The present invention also provides fusion polypeptides including, for example, heterologous fusion polypeptides, e.g., a fusion between an OsMADS polypeptide sequence or fragment thereof and a heterologous polypeptide sequence, e.g., a sequence from a different polypeptide. Such heterologous fusion polypeptides generally exhibit biological properties (such as ligand-binding, catalysis, secretion signals, antigenic determinants, etc.) derived from each of the fused sequences. Fusion partners include, for example, immunoglobulins, beta galactosidase, trpE, protein A, beta lactamase, alpha amylase, alcohol dehydrogenase, yeast alpha mating factor, and various signal and leader sequences which, e.g., can direct the secretion of the polypeptide. Fusion polypeptides can be made, for example, by the expression of recombinant nucleic acids or by chemical synthesis.

Polypeptide sequence determination. The sequence of a polypeptide can be determined by any conventional methods. In order to determine the sequence of a polypeptide, the polypeptide is typically fragmented, the fragments separated, and the sequence of each fragment determined. To obtain fragments of a polypeptide for sequence determination, for example, the polypeptide can be digested with an enzyme such as trypsin, clostripain, or *Staphylococcus* protease, or with chemical agents such as cyanogen bromide, o-iodosobenzoate, hydroxylamine or 2-nitro-5-thiocyanobenzoate. Peptide fragments can be separated, e.g., by reversed-phase high-performance liquid chromatography (HPLC) and analyzed by gas-phase sequencing, for example.

Antibodies

The present invention also encompasses polyclonal and/or monoclonal antibodies capable of specifically binding to any of the polypeptides disclosed herein. Such antibodies can be produced by any conventional method. "Specific" antibodies are capable of distinguishing a given polypeptide from other polypeptides in a sample. Specific antibodies are useful, for example in purifying a polypeptide from a biological sample; in cloning alleles or homologs of a given gene sequence from an expression library; as antibody probes for protein blots and immunoassays; etc.

For the preparation and use of antibodies according to the present invention, including various antibody labelling and immunoassay techniques and applications, *see, e.g.*, Goding, Monoclonal Antibodies: Principles and Practice, 2d ed, Academic Press, New York, 1986; and Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1988. Suitable labels for antibodies include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, magnetic particles and the like.

The invention will be better understood by reference to the following Examples, which are intended to merely illustrate the best mode now known for practicing the invention. The scope of the invention is not to be considered limited thereto, however.

EXAMPLES

EXAMPLE 1: Isolation and Characterization of Three Rice MADS-Box Genes That Control the Timing of Flowering -- *OsMADS6*, *OsMADS7*, and *OsMADS8*

Experimental procedures

Bacterial Strains, Plant Materials, and Plant Transformation. *Escherichia coli* MC1000 (ara, leu, lac, gal, str) was used as the recipient for routine cloning experiments. Rice (*Oryza sativa L.* cv. M201) plants were grown in a growth chamber at 26°C with 10.5-hr day cycle.

Bacterial strains, plant materials, and plant transformation. *Escherichia coli* JM 83 was used as the recipient for routine cloning experiments. *Agrobacterium tumefaciens* LBA4404 (Hoekema et al., Nature 303:179-181 1983) containing the *Ach5* chromosomal background and a disarmed helper-Ti plasmid pAL4404 was used for transformation of tobacco plants (*N. tabacum L.* cv. Xanthi) by the cocultivation method (An et al., in: Plant Molecular Biology Manual, Gelvin and Schilperoort, eds., Kluwer Academic, Dordrecht, Belgium, pp. A3/1-19, 1988). Transgenic tobacco plants were maintained under greenhouse conditions. Rice (*Oryza sativa* L. cv. M201) plants were grown in a growth chamber at 29°C with a 10.5 h day cycle.

Library screening and sequence analysis. cDNA libraries were constructed using λ ZapII vector (Stratagene) and mRNA was prepared from rice flowers at floral primordial stage when the length of the panicles was below 1 cm. Hybridization was performed with 10^5 plaques using a ^{32}P-labeled probe of the *OsMADS1* coding region. The cDNA insert was rescued *in vivo* using an f1 helper phage, R408 (Stratagene). Both strands of the cDNA were sequenced by the dideoxy-nucleotide chain termination method using a double-strand DNA as a template (Sanger et al. Proc. Natl. Acad. Sci. USA 74:5463-5467, 1977). Protein-sequence similarity was analyzed by the IG Suite software package (Intelligenetics Co., Mountain View, CA) and the NCBI non-redundant protein database on the international network.

DNA and RNA blot analyses. Genomic DNA was isolated by the cetyltrimethylammonium bromide (CTAB) method from two-week-old rice seedlings grown hydroponically (Rogers and Bendich, Plant Molecular Biology Manual Kluwer Academic, Dordrecht, Belguim, pp. A6/1-

101988). Eight μg of genomic DNA was digested with the appropriate restriction enzymes, separated on a 0.7% agarose gel, blotted onto a nylon membrane, and hybridized with a ^{32}P-labeled probe for 16 h at 65°C, followed by a wash with a solution containing 2X SSC and 0.5% SDS for 20 min at 65°C, followed by a wash with a solution of 0.1X SSC and 0.1% SDS for 15 min at the same temperature. Total RNA was isolated by the guanidium thiocyanate method (Sambrook et al., 1989). Leaf and root samples were harvested from the two-week-old seedlings. Floral organ samples were obtained by dissecting late vacuolated-stage flowers under a dissecting microscope. Twenty-five μg of total RNA was fractionated on a 1.3% agarose gel as described previously (Sambrook et al., 1989). After RNA transfer onto a nylon membrane, the resulting blot was hybridized in a solution containing 0.5 M $NaPO_4$ (pH 7.2), 1 mM EDTA, 1% BSA, and 7% SDS for 20 h at 60°C (Church and Gilbert, Proc. Natl. Acad. Sci. USA 81:1191-1195, 1984). After hybridization, the blot was washed twice with a solution containing 0.1X SSPE and 0.1% SDS for 5 min at room temperature followed by two washes of the same solution at 60°C for 15 min.

Mapping procedures. An F11 recombinant inbred population consisting of 164 lines derived from a cross between Milyang 23 and Gihobyeo was used to construct a molecular map. Three-week old leaf tissue was harvested from over one hundred seedlings for each F11 line and bulked for DNA extraction. DNA was digested with restriction enzymes (*Bam*HI, *Dra*I, *Eco*RI, *Hind*III, *Eco*RV, *Sca*I, *Xba*I, *Kpn*I) and 8 μg per lane was used to make mapping filters. DNA blotting and hybridization were performed as described above. Linkage analysis was performed using Mapmaker Version 3.0 (Lander et al., Genomics 12:174-181, 1987) on a Macintosh Power PC 8100/80AV. Map units (cM) were derived using the Kosambi function (Kosambi, Ann. Eugen. 12:172-175, 1994).

Results

Isolation of rice cDNA clones encoding MADS box protein. Three cDNA clones were isolated by screening a λ ZapII cDNA library that was prepared from rice floral primordia using *OsMADS1* cDNA as a probe (described above). These clones were designated *OsMADS6*, *OsMADS7*, and *OsMADS8*. DNA sequence analysis showed that these clones are 1180 bp to 1259 bp long and encode putative proteins 248 to 250 amino acid residues long (*OsMADS6*: FIG. 1, SEQ ID NO:2 and SEQ ID NO:3; *OsMADS7*: FIG. 2, SEQ ID NO:4 and SEQ ID NO:5; *OsMADS8*: FIG. 3, SEQ ID NO:6 and SEQ ID NO:7). The 5'-untranslated region of the *OsMADS8* cDNA contains eight repeats of the GGA sequence and the 5'-untranslated region of *OsMADS7* cDNA contains six repeats of the GGT sequence, a so-called microsatellite (Browne and Litt, Nucl. Acids Res. 20:141, 1991; Stalings, Genomics 17:890-891, 1992). Such repeat sequences have been observed in other rice MADS-box genes (Chung et al., Plant Mol. Biol. 26:657-665, 1994).

The MADS-box domain of the cDNA clone is located between the 2nd and 57th amino acids of each protein (FIG. 1 and SEQ ID NO:3; FIG. 2 and SEQ ID NO:5; FIG. 3 and SEQ ID

NO:7). Comparison with other MADS-box genes shows that this region is the most conserved. A second conserved domain, the K box, is located between residues 91 and 156 in *OsMADS6* and between the residues 95 and 160 in both *OsMADS7* and *OsMADS8* (FIGS. 1-3). These genes contain two variable regions, the I-region between the MADS and K boxes, and the C-region downstream of the K box (Purugganan et al., Genetics 140:345-356, 1995). The structure of the proteins encoded by *OsMADS6*, *OsMADS7*, and *OsMADS8* is therefore typical of the plant MADS-box gene family.

Based on the amino-acid sequence similarity of the entire coding region, the OsMADS6, OsMADS7, and OsMADS8 proteins can be grouped into the AGL2 family, which includes AGL2, AGL4 and AGL6 of *Arabidopsis* (Ma et al., Genes Dev. 5:484-495, 1991), ZAG3 and ZAG5 of maize (Mena et al., Plant J. 8:845-854, 1995), FBP2 of petunia (Angenent et al., Plant J. 5:33-44, 1994), TM5 of tomato (Pnueli et al., Plant Cell 6:175-186, 1994), OM1 of orchid (Lu et al., Plant Mol. Biol. 23:901-904, 1993), and OsMADS1 of rice. Among these genes, the OsMADS6 protein was most homologous to ZAG3 (84% homology) and ZAG5 (82% homology), while the OsMADS7 and OsMADS8 proteins were most homologous to OM1 (61% and 65%, respectively) and FBP2 (60% and 64% homology, respectively). OsMADS6, OsMADS7, and OsMADS8 proteins had 50% amino acid sequence homology to OsMADS1.

Alignment of the OsMADS6, OsMADS7, and OsMADS8 proteins with other members of the AGL2 family showed that the MADS-box (FIG. 4A; SEQ ID NOS:1, 3, 5, 7 and 8-15), K-box (FIG. 4B; SEQ ID NOS:1, 3, 5, 7 and 20-27), and C-terminal end regions (FIG. 4C; SEQ ID NOS:1, 3, 5, 7 and 32-39) share significant sequence homologies. The MADS-box region of OsMADS6 is 100% identical to that of ZAG3 and differs from the MADS-box region of OsMADS7 and OsMADS8 in two positions; the 22nd and 50th amino acid serines in OsMADS6 are replaced with alanine and asparagine, respectively, in both OsMADS7 and OsMADS8. The MADS-box sequences of OsMADS6, OsMADS7, and OsMADS8 share at least 89% identity to the MADS-box sequences of other AGL2 proteins. The sequence homology in the K-box region is lower compared to the MADS-box region, but still significant. These regions of OsMADS6, OsMADS7, and OsMADS8 are at least 43% identical to other members of the family, whereas the homology was much lower with distantly related MADS-box proteins such as AG, AP3, and PI. The sequence homology at the C-terminal end was much lower. However, there are two blocks of conserved regions at the end of the proteins, and these AGL2-specific sequences were not found in other MADS-box proteins. In general, MADS-box proteins include (beginning from the amino-terminus): a MADS-box region, an I region, a K-box region, a C-terminal region, and a C-terminal end region.

RNA blot analysis. There are a large number of MADS-box genes in the rice genome. Genomic DNA blot analyses were conducted to identify a region that would not cross hybridize with other MADS box genes. Rice genomic DNA was digested with *Eco*RI, *Hind*III, or *Pst*I,

fractionated on a 0.8% agarose gel, and hybridized with the gene specific probes located at the 3' region of each cDNA. The 300 bp *Pst*I-*Eco*RI fragment, which is located at the C-terminal region of *OsMADS6*, hybridized to single DNA fragments. Likewise, the 280 bp *Pst*I-*Eco*RI fragment of *OsMADS7* and 220 bp *Nhe*I-*Eco*RI fragment of *OsMADS8* were shown to be gene-specific regions.

RNA blot analyses of the *OsMADS6*, *OsMADS7*, and *OsMADS8* transcripts in rice were conducted. Ten μg of total RNA isolated from roots and leaves of two-week-old seedlings, and paleas/lemmas, anthers, and carpels of late vacuolated-stage flowers were hybridized with the gene-specific probes.

The *OsMADS6* transcript was detectable primarily in carpels and also weakly in palea and lemma of late vacuolated pollen-stage flowers. However, the transcript was not detectable in anthers or vegetative organs. This expression pattern is similar to that of *OsMADS1*. Spatial expression patterns of the *OsMADS7* and *OsMADS8* clones were different from that of *OsMADS6*. Transcripts of both clones were detectable primarily in carpels and also weakly in anthers. This expression pattern is similar to those of *OsMADS3* and *OsMADS4* (Chung et al., Plant Science 109:45-56, 1995; Kang et al., Plant Mol. Biol. 29:1-10, 1995).

The temporal expression pattern of *OsMADS* genes during flower development was also examined. Twenty-five μg of total RNA isolated from rice flowers at three different developmental stages were used for detection of the *OsMADS* gene expression. The stages examined were: young flowers at the panicle size (1 to 5 cm); flowers at the early vacuolated pollen stage; and flowers at the late vacuolated pollen stage. Ethidium bromide staining of 25S and 17S rRNAs were shown to demonstrate equal amounts of RNA loading. During flower development the *OsMADS6* and *OsMADS7* genes were strongly expressed at the young flower stage and expression gradually decreased as the flower further developed to the mature flower stage. The expression of *OsMADS8* was weak at the young flower stage and expression gradually increased as the flower developed.

Chromosomal mapping of the OsMADS genes. An F_{11} recombinant inbred population of rice was used to locate the *OsMADS* genes on a genetic map. C-terminal DNA fragments that were shown to be unique to each *OsMADS* gene were used. These experiments revealed that *OsMADS6* is located on the long arm of chromosome 2, *OsMADS7* on the long arm of chromosome 8, and *OsMADS8* on the long arm of chromosome 9 (FIG. 5). FIG. 5 indicates the location of two other MADS genes, *OsMADS2* and *OsMADS3*. *OsMADS2* is a member of *GLOBOSA* family and is located on the long arm of chromosome 1 (FIG. 5). *OsMADS3* is a rice homolog of *Arabidopsis AGAMOUS* and is located on the short arm of chromosome 1.

Ectopic expression. The functional roles of the three rice MADS genes were studied using tobacco plants as a heterologous expression system. The cDNA clones were placed under the control of the CaMV 35S promoter and transcript 7 terminator using the binary vector pGA748 (An et al., in: Plant Molecular Biology Manual, Gelvin and Schilperoort, eds., Kluwer Academic, Dordrecht, Belgium, pp. A3/1-19, 1988). The chimeric molecules were transferred to tobacco

plants using a kanamycin-resistance marker and an *Agrobacterium*-mediated Ti plasmid vector system. Ten independent T1 transgenic plants were regenerated to avoid any artifacts. Some of the primary transgenic plants were shorter and bloomed earlier than control plants, which were transformed with the Ti plasmid vector alone, while others showed normal growth.

RNA-blot analyses of transgenic plants expressing the *OsMADS6*, *OsMADS7*, and *OsMADS8* transcripts were performed in order to investigate the expression level of the transgenes. Ten μg of total RNAs isolated from young leaves of the tobacco transgenic plants were hybridized with gene-specific probes. In order to minimize the variation due to the stage of development, young leaves at anthesis of the first flower were used for RNA isolation. It was found that plants showing the early flowering phenotype expressed higher levels of the transgene compared with transgenic plants exhibiting a weak or no early flowering phenotype. The transgenic lines OsMADS6-2, -4, -5, and -7 accumulated higher levels of the transgene transcript and flowered earlier. Transgenic lines OsMADS7-5, -9, and -10, and OsMADS8-4 and -5 accumulated higher levels of the transgene transcript and flowered earlier.

Transgenic lines (T2 generation) that expressed *OsMADS6*, *OsMADS7*, and *OsMADS8* and displayed the most severe phenotypes were selected to examine the inheritance of the characteristics. The results showed that the early flowering phenotypes was co-inherited with the kanamycin resistance gene to the next generation. The transgenic plant line OsMADS6-7 flowered an average of 10 days earlier that control plants and was 30 cm shorter than controls. Similarly, both OsMADS7-10 and OsMADS8-5 flowered an average of nine days earlier than control plants and were significantly shorter than wild-type control plants.

Discussion

The three additional rice MADS-box genes that were isolated are probably involved in controlling the timing of flowering. The deduced amino acid sequences of the gene products showed a high homology to the AGL2 family proteins. The homology was extensive, covering the entire protein. It was observed that the AGL2 family of proteins could be further divided into several subgroups depending on the protein sequence similarity in the K box and the two variable regions (Theißen and Saedler, Curr. Opinion in Genet. and Dev. 5:628-639, 1995). Our results (see FIG. 4A-C) show that *OsMADS6* belongs to the AGL6 subfamily and *OsMADS7* and *OsMADS8* both belong to the FBP2 subfamily.

The sequence identity of these genes suggest that they share similar biological function. Using the co-suppression approach (Angenent et al., Plant J. 5:33-44, 1994), it was found that suppression of FBP2 expression in petunia flowers resulted in aberrant flowers with modified whorl two, three, and four organs. The flower possessed a green corolla, petaloid stamens, and dramatically altered carpel structure. Therefore, FBP2 is apparently involved in the determination of the central parts of the generative meristem. Using an antisense RNA approach, TM5 has been

observed to have similar effects on the development of the three whorls. It is known that transgenic plants overexpressing *OsMADS1* exhibit early flowering and dwarf phenotypes, indicating that *OsMADS1* is involved in controlling the timing of flowering without altering the morphology of the floral organs. These observations suggest that the FBP2 and TM5 genes function differently than the *OsMADS1* gene. Interestingly, the length of the OsMADS6, OsMADS7, and OsMADS8 proteins is similar to the OsMADS1 and AP1 proteins, but much longer than the FBP2 and TM5 proteins. Therefore, it is possible that the additional amino acid sequences encoded by the *OsMADS6-8* genes are responsible for controlling the timing of flowering.

RNA blot analyses showed that the *OsMADS6*, *OsMADS7*, and *OsMADS8* genes were expressed specifically in flowers, which coincides with the expression of genes of the AGL2 family. This indicates that the genes of the AGL2 family function primarily during the flower development. The expression of the *OsMADS* genes started at the early stage of the flower development and extended into the later stages of flower development, indicating that the *OsMADS6-8* genes play critical roles during the early stages and continue to function as the flower further develops. Such expression patterns were also observed from other AGL2 members, including AGL2, AGL4, FBP2, TM5 (Angenent et al., Plant Cell 4:983-993, 1992; Ma et al., Genes Dev. 5:484-495, 1991; Pnueli et al., Plant J. 1:255-266, 1991), and *OsMADS1*. However, not all members of the AGL2 family are expressed at early stages of development. The OM1 transcript is detectable only after flower organs have fully developed (Lu et al., Plant Mol. Biol. 23:901-904, 1993). In mature flowers, the *OsMADS6* gene was preferentially expressed in the carpels and palea/lemma. Similar expression patterns have been found in *OsMADS1*, AP1, and SQUA, suggesting a possibility that they belong to a functionally similar group. The FBP2 and TM5 genes are expressed in the whorls 2, 3 and 4 (Pnueli et al., Plant Cell 6:175-186, 1994; Angenent et al., Plant Cell 4:983-993, 1992). Unlike most dicots, rice flowers contain a single perianth, the palea/lemma, which more closely resembles a sepal than a petal. The palea/lemma contains chlorophyll and remains attached to mature seeds. Therefore, expression of FBP2 homologs in dicots is expected to be restricted in sepals and petals. The *OsMADS7* and *OsMADS8* genes were expressed in the inner two whorls, coinciding with the expected expression pattern.

OsMADS6, *OsMADS7*, and *OsMADS8* mapped to rice chromosomes 2, 8, and 9, respectively. The EF-1 gene, which controls the timing of flowering in rice, is located on chromosome 10, and the Se genes, which determine photoperiod sensitivity, are located on chromosomes 6 or 7 (Khush and Kinoshita, in *Rice Biotechnology*, Khush and Toennesson, eds., C.A.B. International and International Rice Research Institute, pp. 93-106, 1991). Therefore, it is evident that none of the early flowering MADS-box genes are linked to previously mapped markers that are involved in controlling the timing of flowering. The relationship of *OsMADS6*, *OsMADS7*,

and *OsMADS8* to other genes involved in the timing of flowering, such as E-1, E-2, E-3, *lf*-1 and *lf*-2, can be resolved when these genes are mapped.

To elucidate the functions of the rice MADS-box genes, we have generated transgenic tobacco plants that express a chimeric fusion between the CaMV 35S promoter and an OSMADS cDNA. *OsMADS6, OsMADS7,* and *OsMADS8* genes caused early flowering and dwarf phenotypes when strongly expressed in transgenic plants.

This invention has been detailed both by example and by direct description. It should be apparent that one having ordinary skill in the relevant art would be able to surmise equivalents to the invention as described in the claims which follow but which would be within the spirit of the foregoing description. Those equivalents are to be included within the scope of this invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANTS: An, Gynheung

(ii) TITLE OF INVENTION: GENES CONTROLLING FLORAL DEVELOPMENT AND APICAL DOMINANCE IN PLANTS

(iii) NUMBER OF SEQUENCES: 43

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Klarquist Sparkman Campbell Leigh & Whinston

(B) STREET: One World Trade Center
121 S.W. Salmon Street
Suite 1600

(C) CITY: Portland

(D) STATE: Oregon

(E) COUNTRY: United States of America

(F) ZIP: 97204

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Disk, 3-1/2 inch

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: MS DOS

(D) SOFTWARE: WordPerfect 7.0 for Windows

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE: June 2, 1998

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: U.S. 08/867,087

(B) FILING DATE: June 2, 1997

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Stephens, Donald L., Jr.

(B) REGISTRATION NUMBER: 34,022

(C) REFERENCE/DOCKET NUMBER: 4630-50302

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (503) 226-7391

(B) TELEFAX: (503) 228-9446

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 257 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Gly Arg Gly Lys Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Ser
5 10 15

Arg Gln Val Thr Phe Ala Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala
20 25 30

Tyr Glu Leu Ser Leu Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe
35 40 45

Ser Gly Arg Gly Arg Leu Phe Glu Phe Ser Ser Ser Ser Cys Met Tyr
50 55 60

Lys Thr Leu Glu Arg Tyr Arg Ser Cys Asn Tyr Asn Ser Gln Asp Ala
65 70 75 80

Ala Ala Pro Glu Asn Glu Ile Asn Tyr Gln Glu Tyr Leu Lys Leu Lys
85 90 95

Thr Arg Val Glu Phe Leu Gln Thr Thr Gln Arg Asn Ile Leu Gly Glu
100 105 110

Leu Ser Met Asp Leu Gly Pro Lys Glu Leu Glu Gln Leu Glu Asn Gln
115 120 125

Ile Glu Val Ser Leu Lys Gln Ile Arg Ser Arg Lys Asn Gln Ala Leu
130 135 140

Leu Asp Gln Leu Phe Asp Leu Lys Ser Lys Glu Gln Gln Leu Gln Asp
145 150 155 160

Leu Asn Lys Asp Leu Arg Lys Lys Leu Gln Glu Thr Ser Ala Glu Asn
165 170 175

Val Leu His Met Ser Trp Gln Asp Gly Gly Gly His Ser Gly Ser Ser
180 185 190

Thr Val Leu Ala Asp Gln Pro His His His Gln Gly Leu Leu His Pro
195 200 205

His Pro Asp Gln Gly Asp His Ser Leu Gln Ile Gly Tyr His His Pro
210 215 220

His Ala His His His Gln Ala Tyr Met Asp His Leu Ser Asn Glu Ala
225 230 235 240

Ala Asp Met Val Ala His His Pro Asn Glu His Ile Pro Ser Gly Trp
245 250 255

Ile

(3) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1060 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double stranded

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
TACCCGCGGG AATCGTTCGA TCGATCGGGC GAG                                    33

ATG GGG AGG GGA AGA GTT GAG CTG AAG CGC ATC GAG AAC AAG ATC AAC         81
Met Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn
                5                  10                  15

AGG CAG GTC ACC TTC TCC AAG CGC CGC AAC GGC CTC CTC AAG AAG GCC        129
Arg Gln Val Thr Phe Ser Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala
            20                  25                  30

TAC GAG CTG TCC GTT CTC TGC GAC GCC GAG GTC GCG CTC ATC ATC TTC        177
Tyr Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe
        35                  40                  45

TCC AGC CGC GGC AAG CTC TAC GAG TTC GGC AGC GCC GGC ATA ACA AAG        225
Ser Lys Ser Arg Gly Leu Tyr Glu Phe Gly Ser Ala Gly Ile Thr Lys
    50                  55                  60

ACT TTA GAA AGG TAC CAA CAT TGT TGC TAC AAT GCT CAA GAT TCC AAC        273
Thr Leu Glu Arg Tyr Gln His Cys Cys Tyr Asn Ala Gln Asp Ser Asn
65                  70                  75                  80

AAT GCA CTT TCT GAA ACT CAG AGT TGG TAC CAT GAA ATG TCA AAG TTG        321
Asn Ala Leu Ser Glu Thr Gln Ser Trp Tyr His Glu Met Ser Lys Leu
                85                  90                  95

AAA GCA AAA TTT GAA GCT TTG CAG CGC ACT CAA AGG CAC TTG CTT GGG        369
Lys Ala Lys Phe Glu Ala Leu Gln Arg Thr Gln Arg His Leu Leu Gly
            100                 105                 110

GAG GAT CTT GGA CCA CTC AGC GTC AAA GAA TTG CAG CAG CTG GAG AAA        417
Glu Asp Leu Gly Pro Leu Ser Val Lys Glu Leu Gln Gln Leu Glu Lys
        115                 120                 125

CAG CTT GAA TGT GCA CTA TCA CAG GCG AGA CAG AGA AAG ACG CAA CTG        465
Gln Leu Glu Cys Ala Leu Ser Gln Ala Arg Gln Arg Lys Thr Gln Leu
    130                 135                 140

ATG ATG GAA CAG GTG GAG GAA CTT CGC AGA AAG GAG CGT CAG CTG GGT        513
Met Met Glu Gln Val Glu Glu Leu Arg Arg Lys Glu Arg Gln Leu Gly
145                 150                 155                 160

GAA ATT AAT AGG CAA CTC AAG CAC AAG CTC GAG GTT GAA GGT TCC ACC        561
Glu Ile Asn Arg Gln Leu Lys His Lys Leu Glu Val Glu Gly Ser Thr
                165                 170                 175

AGC AAC TAC AGA GCC ATG CAG CAA GCC TCC TGG GCT CAG GGC GCC GTG        609
Ser Asn Tyr Arg Ala Met Gln Gln Ala Ser Trp Ala Gln Gly Ala Val
            180                 185                 190

GTG GAG AAT GGC GCC GCA TAC GTG CAG CCG CCG CCA CAC TCC GCG GCC        657
Val Glu Asn Gly Ala Ala Tyr Val Gln Pro Pro Pro His Ser Ala Ala
        195                 200                 205

ATG GAC TCT GAA CCC ACC TTG CAA ATT GGG TAT CCT CAT CAA TTT GTG        705
Met Asp Ser Glu Pro Thr Leu Gln Ile Gly Tyr Pro His Gln Phe Val
    210                 215                 220

CCT GCT GAA GCA AAC ACT ATT CAG AGG AGC ACT GCC CCT GCA GGT GCA        753
Pro Ala Glu Ala Asn Thr Ile Gln Arg Ser Thr Ala Pro Ala Gly Ala
225                 230                 235                 240

GAG AAC AAC TTC ATG CTG GGA TGG GTT CTT TGA                            786
Glu Asn Asn Phe Met Leu Gly Trp Val Leu
                245                 250

GCTAAGCAGC CATCGATCAG CTGTCAGAAG TTGGAGCTAA TAATAAAAGG GATGTGGAGT      846
```

GGGCTACATG TATCTCGGAT CTCTCTGCGA GCCACCTAAT GGTCTTGCGT GGCCCTTTAA 906

TCTGTATGTT TTTGTGTGTA AGCTACTGCT AGCTGTTTGC ACCTTCTGCG TCCGTGGTTG 1026

TGTTTCCGTG CTACCTTTTT ATGTTTTGAT TTGGATCTTG TTTGAAAATA ATCTTACCAG 1043

CTTTGGGTAA ACTGTTT 1060

(4) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 250 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn
5 10 15

Arg Gln Val Thr Phe Ser Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala
20 25 30

Tyr Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe
35 40 45

Ser Lys Ser Arg Gly Leu Tyr Glu Phe Gly Ser Ala Gly Ile Thr Lys
50 55 60

Thr Leu Glu Arg Tyr Gln His Cys Cys Tyr Asn Ala Gln Asp Ser Asn
65 70 75 80

Asn Ala Leu Ser Glu Thr Gln Ser Trp Tyr His Glu Met Ser Lys Leu
85 90 95

Lys Ala Lys Phe Glu Ala Leu Gln Arg Thr Gln Arg His Leu Leu Gly
100 105 110

Glu Asp Leu Gly Pro Leu Ser Val Lys Glu Leu Gln Gln Leu Glu Lys
115 120 125

Gln Leu Glu Cys Ala Leu Ser Gln Ala Arg Gln Arg Lys Thr Gln Leu
130 135 140

Met Met Glu Gln Val Glu Glu Leu Arg Arg Lys Glu Arg Gln Leu Gly
145 150 155 160

Glu Ile Asn Arg Gln Leu Lys His Lys Leu Glu Val Glu Gly Ser Thr
165 170 175

Ser Asn Tyr Arg Ala Met Gln Gln Ala Ser Trp Ala Gln Gly Ala Val
180 185 190

Val Glu Asn Gly Ala Ala Tyr Val Gln Pro Pro Pro His Ser Ala Ala
195 200 205

Met Asp Ser Glu Pro Thr Leu Gln Ile Gly Tyr Pro His Gln Phe Val
210 215 220

Pro Ala Glu Ala Asn Thr Ile Gln Arg Ser Thr Ala Pro Ala Gly Ala
225 230 235 240

Glu Asn Asn Phe Met Leu Gly Trp Val Leu
245 250

(5) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1059 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double stranded

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TATCCCCTTC CTCCAGGTGG CTTGTTTCTT GCAGTGGTGG TGGTGGTGGT GGTGAGATCT 60

AGCTTGGTTG GTTGGTGGCA GCTGGAGATC GATCGGG 97

ATG GGG AGG GGG CGG GTG GAG CTG AAG AGG ATC GAG AAC AAG ATC AAC 145
Met Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn
5 10 15

CGG AAG GTG ACG TTC GCC AAG AGG AGG AAT GGC CTG CTC AAG AAG GCG 193
Arg Lys Val Thr Phe Ala Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala
20 25 30

TAC GAG CTC TCC GTC CTC TGC GAC GCC GAG GTC GCC CTC ATC ATC TTC 241
Tyr Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe
35 40 45

TCC AAC CGC GGC AAG CTC TAC GAG TTC TGC AGC ACC CAG AGC ATG ACT 289
Ser Asn Arg Gly Lys Leu Tyr Glu Phe Cys Ser Thr Gln Ser Met Thr
50 55 60

AAA ACG CTT GAG AAG TAT CAG AAA TGC AGT TAC GCA GGA CCC GAA ACA 337
Lys Thr Leu Glu Lys Tyr Gln Lys Cys Ser Tyr Ala Gly Pro Glu Thr
65 70 75 80

GCT GTC CAA AAT AGA GAA AGT GAG CAA TTG AAA GCT AGC CGC AAT GAA 385
Ala Val Gln Asn Arg Glu Ser Glu Gln Leu Lys Ala Ser Arg Asn Glu
85 90 95

TAC CTC AAA CTG AAG GCA AGG GTT GAA AAT TTA CAA CGG ACT CAA AGA 433
Tyr Leu Lys Leu Lys Ala Arg Val Glu Asn Leu Gln Arg Thr Gln Arg
100 105 110

AAT TTG CTG GGT CCA GAT CTT GAT TCA TTA GGC ATA AAA GAG CTC GAG 481
Asn Leu Leu Gly Pro Asp Leu Asp Ser Leu Gly Ile Lys Glu Leu Glu
115 120 125

AGC CTA GAG AAG CAG CTT GAT TCA TCC CTG AAG CAC GTC AGA ACT ACA 529
Ser Leu Glu Lys Gln Leu Asp Ser Ser Leu Lys His Val Arg Thr Thr
130 135 140

AGG ACA AAA CAT CTG GTC GAC CAA CTG ACG GAG CTT CAG AGA AAG GAA 577
Arg Thr Lys His Leu Val Asp Gln Leu Thr Glu Leu Gln Arg Lys Glu
145 150 155 160

CAA ATG GTT TCT GAA GCA AAT AGA TGC CTT AGG AGA AAA CTG GAG GAA 625
Gln Met Val Ser Glu Ala Asn Arg Cys Leu Arg Arg Lys Leu Glu Glu
165 170 175

AGC AAC CAT GTT CGC GGG CAG CAA GTG TGG GAG CAG GGC TGC AAC TTA 673
Ser Asn His Val Arg Gly Gln Gln Val Trp Glu Gln Gly Cys Asn Leu
180 185 190

ATT GGC TAT GAA CGT CAG CCT GAA GTG CAG CAG CCT CTT CAC GGC GGC 721
Ile Gly Tyr Glu Arg Gln Pro Glu Val Gln Gln Pro Leu His Gly Gly
195 200 205

AAT GGG TTC TTC CAT CCA CTT GAT GCT GCT GGT GAA CCC ACC CTT CAG 769
Asn Gly Phe Phe His Pro Leu Asp Ala Ala Gly Glu Pro Thr Leu Gln

```
     210                 215                 220

ATT GGG TAC CCT GCA GAG CAT CAT GAG GCG ATG AAC AGT GCG TGC ATG     817
Ile Gly Tyr Pro Ala Glu His His Glu Ala Met Asn Ser Ala Cys Met
225                 230                 235                 240

AAC ACC TAC ATG CCC CCA TGG CTA CCA TGA                          847
Asn Thr Tyr Met Pro Pro Trp Leu Pro
                245

TGATGACGGG ACAATGAATT ACGAAATAAC AAGGATATGT GGCATATATG TGGTGCCGCA  907

TACATGCATG TATCATGGCT AGCTACTTAA TTGGAGTGAT GGATTTGAAC TAGTTTCGTA  967

TGTAGCCTGT TTGTGTGTAA CTTGTGTGAG ATACTACCTT AAAAACTATC GGTGTCTGTT 1027

GAACATATTC TGCGATCAAC TTTAAGCGTA TT                               1059
```

(6) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 249 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Met Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn
                5                  10                  15

Arg Lys Val Thr Phe Ala Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala
            20                  25                  30

Tyr Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe
        35                  40                  45

Ser Asn Arg Gly Lys Leu Tyr Glu Phe Cys Ser Thr Gln Ser Met Thr
    50                  55                  60

Lys Thr Leu Glu Lys Tyr Gln Lys Cys Ser Tyr Ala Gly Pro Glu Thr
65                  70                  75                  80

Ala Val Gln Asn Arg Glu Ser Glu Gln Leu Lys Ala Ser Arg Asn Glu
                85                  90                  95

Tyr Leu Lys Leu Lys Ala Arg Val Glu Asn Leu Gln Arg Thr Gln Arg
            100                 105                 110

Asn Leu Leu Gly Pro Asp Leu Asp Ser Leu Gly Ile Lys Glu Leu Glu
        115                 120                 125

Ser Leu Glu Lys Gln Leu Asp Ser Ser Leu Lys His Val Arg Thr Thr
    130                 135                 140

Arg Thr Lys His Leu Val Asp Gln Leu Thr Glu Leu Gln Arg Lys Glu
145                 150                 155                 160

Gln Met Val Ser Glu Ala Asn Arg Cys Leu Arg Arg Lys Leu Glu Glu
                165                 170                 175

Ser Asn His Val Arg Gly Gln Gln Val Trp Glu Gln Gly Cys Asn Leu
            180                 185                 190

Ile Gly Tyr Glu Arg Gln Pro Glu Val Gln Gln Pro Leu His Gly Gly
        195                 200                 205

Asn Gly Phe Phe His Pro Leu Asp Ala Ala Gly Glu Pro Thr Leu Gln
```

```
     210                 215                 220

Ile Gly Tyr Pro Ala Glu His His Glu Ala Met Asn Ser Ala Cys Met
225                 230                 235                 240

Asn Thr Tyr Met Pro Pro Trp Leu Pro
                245
```

(7) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1180 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double stranded

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
TGCTTTCCCC TCTCTTCCGC TTCGCGAGAT TGGTTGATTC ATCTCGCGAT TGATCGAGCT    60

CGAGCGGCGG TGAGGTGAGG TGGAGGAGGA GGAGGAGGAG GAGATCGGG               109

ATG GGG AGA GGG AGG GTG GAG CTG AAG AGG ATC GAG AAC AAG ATC AAC     157
Met Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn
                5                   10                  15

AGG CAG GTG ACG TTC GCG AAG CGG AGG AAT GGG CTG CTC AAG AAG GCG     205
Arg Gln Val Thr Phe Ala Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala
            20                  25                  30

TAC GAG CTC TCC GTG CTC TGC GAC GCC GAG GTC GCC CTC ATC ATC TTC     253
Tyr Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe
        35                  40                  45

TCC AAC CGC GGC AAG CTC TAC GAG TTC TGC AGC GGC CAA AGC ATG ACC     301
Ser Asn Arg Gly Lys Leu Tyr Glu Phe Cys Ser Gly Gln Ser Met Thr
    50                  55                  60

AGA ACT TTG GAA AGA TAC CAA AAA TTC AGT TAT GGT GGG CCA GAT ACT     349
Arg Thr Leu Glu Arg Tyr Gln Lys Phe Ser Tyr Gly Gly Pro Asp Thr
65                  70                  75                  80

GCA ATA CAG AAC AAG GAA AAT GAG TTA GTG CAA AGC AGC CGC AAT GAG     397
Ala Ile Gln Asn Lys Glu Asn Glu Leu Val Gln Ser Ser Arg Asn Glu
                85                  90                  95

TAC CTC AAA CTG AAG GCA CGG GTG GAA AAT TTA CAG AGG ACC CAA AGG     445
Tyr Leu Lys Leu Lys Ala Arg Val Glu Asn Leu Gln Arg Thr Gln Arg
            100                 105                 110

AAT CTT CTT GGT GAA GAT CTT GGG ACA CTT GGC ATA AAA GAG CTA GAG     493
Asn Leu Leu Gly Glu Asp Leu Gly Thr Leu Gly Ile Lys Glu Leu Glu
        115                 120                 125

CAG CTT GAG AAA CAA CTT GAT TCA TCC TTG AGG CAC ATT AGA TCC ACA     541
Gln Leu Glu Lys Gln Leu Asp Ser Ser Leu Arg His Ile Arg Ser Thr
    130                 135                 140

AGG ACA CAG CAT ATG CTT GAT CAG CTC ACT GAT CTC CAG AGG AGG GAA     589
Arg Thr Gln His Met Leu Asp Gln Leu Thr Asp Leu Gln Arg Arg Glu
145                 150                 155                 160

CAA ATG TTG TGT GAA GCA AAT AAG TGC CTC AGA AGA AAA CTG GAG GAG     637
Gln Met Leu Cys Glu Ala Asn Lys Cys Leu Arg Arg Lys Leu Glu Glu
                165                 170                 175
```

AGC AAC CAG TTG CAT GGA CAA GTG TGG GAG CAC GGC GCC ACC CTA CTC 685
Ser Asn Gln Leu His Gly Gln Val Trp Glu His Gly Ala Thr Leu Leu
180 185 190

GGC TAC GAG CGG CAG TCG CCT CAT GCC GTC CAG CAG GTG CCA CCG CAC 733
Gly Tyr Glu Arg Gln Ser Pro His Ala Val Gln Gln Val Pro Pro His
195 200 205

GGT GGC AAC GGA TTC TTC CAT TCC CTG GAA GCT GCC GCC GAG CCC ACC 781
Gly Gly Asn Gly Phe Phe His Ser Leu Glu Ala Ala Ala Glu Pro Thr
210 215 220

TTG CAG ATC GGG TTT ACT CCA GAG CAG ATG AAC AAC TCA TGC GTG ACT 829
Leu Gln Ile Gly Phe Thr Pro Glu Gln Met Asn Asn Ser Cys Val Thr
225 230 235 240

GCC TTC ATG CCG ACA TGG CTA CCC TGA 856
Ala Phe Met Pro Thr Trp Leu Pro
245

ACTCCTGAAG GCCGATGCGA CAACCAATAA AAACGGATGT GACGACACAG ATCAAGTCGC 916

ACCATTAGAT TGATCTTCTC CTACAAGAGT GAGACTAGTA ATTCCGCGTT TGTGTGCTAG 976

CGTGTTGAAA CTTTTCTGAT GTGATGCACG CACTTTTAAT TATTATTAAG CGTTCAAGGA 1036

CTAGTATGTG GTATAAAAGC CCGTACGTGA CAGCCTATGG TTATATGCTG CGCAAAAACT 1096

ACGTATGGTA CAGTGCAGTG CCTGTACATT TCATAATTTG CGGGTAAAGT TTATTGACTA 1156

TATATCCAGT GTGTCAAATA TAAT 1180

(8) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 248 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn
5 10 15

Arg Gln Val Thr Phe Ala Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala
20 25 30

Tyr Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe
35 40 45

Ser Asn Arg Gly Lys Leu Tyr Glu Phe Cys Ser Gly Gln Ser Met Thr
50 55 60

Arg Thr Leu Glu Arg Tyr Gln Lys Phe Ser Tyr Gly Gly Pro Asp Thr
65 70 75 80

Ala Ile Gln Asn Lys Glu Asn Glu Leu Val Gln Ser Ser Arg Asn Glu
85 90 95

Tyr Leu Lys Leu Lys Ala Arg Val Glu Asn Leu Gln Arg Thr Gln Arg
100 105 110

Asn Leu Leu Gly Glu Asp Leu Gly Thr Leu Gly Ile Lys Glu Leu Glu
115 120 125

Gln Leu Glu Lys Gln Leu Asp Ser Ser Leu Arg His Ile Arg Ser Thr
130 135 140

Arg Thr Gln His Met Leu Asp Gln Leu Thr Asp Leu Gln Arg Arg Glu
145 150 155 160

Gln Met Leu Cys Glu Ala Asn Lys Cys Leu Arg Arg Lys Leu Glu Glu
165 170 175

Ser Asn Gln Leu His Gly Gln Val Trp Glu His Gly Ala Thr Leu Leu
180 185 190

Gly Tyr Glu Arg Gln Ser Pro His Ala Val Gln Gln Val Pro Pro His
195 200 205

Gly Gly Asn Gly Phe Phe His Ser Leu Glu Ala Ala Ala Glu Pro Thr
210 215 220

Leu Gln Ile Gly Phe Thr Pro Glu Gln Met Asn Asn Ser Cys Val Thr
225 230 235 240

Ala Phe Met Pro Thr Trp Leu Pro
245

(9) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn Arg
5 10 15

Gln Val Thr Phe Ser Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala Tyr
20 25 30

Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe Ser
35 40 45

Ser Arg Gly Lys Leu Tyr Glu Phe
50 55

(10) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn Arg
5 10 15

Gln Val Thr Phe Ser Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala Tyr
20 25 30

Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe Ser
35 40 45

Gly Arg Gly Lys Leu Tyr Glu Phe
50 55

(11) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Gly Arg Gly Arg Val Glu Met Lys Arg Ile Glu Asn Lys Ile Asn Arg
5 10 15

Gln Val Thr Phe Ser Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala Tyr
20 25 30

Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe Ser
35 40 45

Ser Arg Gly Lys Leu Tyr Glu Phe
50 55

(12) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn Arg
5 10 15

Gln Val Thr Phe Ala Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala Tyr
20 25 30

Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe Ser
35 40 45

Asn Arg Gly Lys Leu Tyr Glu Phe
50 55

(13) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Gly Lys Ile Asn Arg
5 10 15

Gln Val Thr Phe Ala Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala Tyr
20 25 30

Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe Ser
35 40 45

Asn Arg Gly Lys Leu Tyr Glu Phe
50 55

(14) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Gly Arg Gly Arg Val Glu Leu Lys Met Ile Glu Asn Lys Ile Asn Arg
5 10 15

Gln Val Thr Phe Ala Lys Arg Arg Lys Arg Leu Leu Lys Lys Ala Tyr
20 25 30

Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe Ser
35 40 45

Asn Arg Gly Lys Leu Tyr Glu Phe
50 55

(15) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn Arg
5 10 15

Gln Val Thr Phe Ala Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala Tyr
20 25 30

Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe Ser
35 40 45

Asn Arg Gly Lys Leu Tyr Glu Phe
50 55

(16) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn Arg
5 10 15

Gln Val Thr Phe Ala Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala Tyr
20 25 30

Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ser Leu Ile Val Phe Ser
35 40 45

Asn Arg Gly Lys Leu Tyr Glu Phe

50 55

(17) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Gly Arg Gly Arg Val Gln Leu Lys Arg Ile Glu Asn Lys Ile Asn Arg
5 10 15

Gln Val Thr Phe Ser Lys Arg Arg Ala Gly Leu Leu Lys Lys Ala His
20 25 30

Glu Ile Ser Val Leu Cys Asp Ala Glu Val Ala Leu Val Val Phe Ser
35 40 45

His Lys Gly Lys Leu Phe Glu Tyr
50 55

(18) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Gly Arg Gly Lys Ile Glu Ile Lys Arg Ile Glu Asn Thr Thr Asn Arg
5 10 15

Gln Val Thr Phe Cys Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala Tyr
20 25 30

Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Val Phe Ser
35 40 45

Ser Arg Gly Arg Leu Tyr Glu Tyr
50 55

(19) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Ala Arg Gly Lys Ile Gln Ile Lys Arg Ile Glu Asn Gln Thr Asn Arg
5 10 15

Gln Val Thr Tyr Ser Lys Arg Arg Asn Gly Leu Phe Lys Lys Ala His
20 25 30

Glu Leu Thr Val Leu Cys Asp Ala Arg Val Ser Ile Ile Met Phe Ser
35 40 45

Ser Ser Asn Lys Leu His Glu Tyr
50 55

(20) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Gly Arg Gly Lys Ile Glu Ile Lys Arg Ile Glu Asn Ala Asn Asn Arg
5 10 15

Val Val Thr Phe Ser Lys Arg Arg Asn Gly Leu Val Lys Lys Ala Lys
20 25 30

Glu Ile Thr Val Leu Cys Asp Ala Lys Val Ala Leu Ile Ile Phe Ala
35 40 45

Ser Asn Gly Lys Met Ile Asp Tyr
50 55

(21) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Gln Glu Met Ser Lys Leu Arg Ala Lys Phe Glu Ala Leu Gln Arg Thr
5 10 15

Gln Arg His Leu Leu Gly Glu Glu Leu Gly Pro Leu Ser Val Lys Glu
20 25 30

Leu Gln Gln Leu Glu Lys Gln Leu Glu Cys Ala Leu Ser Gln Ala Arg
35 40 45

Gln Arg Lys Thr Gln Leu Met Met Glu Gln Val Glu Glu Leu Arg Arg
50 55 60

Lys
65

(22) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Gln Glu Met Ser Lys Leu Arg Ala Lys Phe Glu Ala Leu Gln Arg Thr
5 10 15

Gln Arg His Leu Leu Gly Glu Asp Leu Gly Pro Leu Ser Val Lys Glu

20 25 30

Leu Gln Gln Leu Glu Lys Gln Leu Glu Cys Ala Leu Ser Gln Ala Arg
35 40 45

Gln Arg Lys Thr Gln Val Met Met Glu Gln Val Glu Glu Leu Arg Arg
50 55 60

Thr
65

(23) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Gln Glu Val Thr Lys Leu Lys Ser Lys Tyr Glu Ser Leu Val Arg Thr
5 10 15

Asn Arg Asn Leu Leu Gly Glu Asp Leu Gly Glu Met Gly Val Lys Glu
20 25 30

Leu Gln Ala Leu Glu Arg Gln Leu Glu Ala Ala Leu Thr Ala Thr Arg
35 40 45

Gln Arg Lys Thr Gln Val Met Met Glu Glu Met Glu Asp Leu Arg Lys
50 55 60

Lys
65

(24) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Gln Glu Tyr Leu Lys Leu Lys Ala Arg Tyr Glu Ala Leu Gln Arg Ser
5 10 15

Gln Arg Asn Leu Leu Gly Glu Asp Leu Gly Pro Leu Asn Ser Lys Glu
20 25 30

Leu Glu Ser Leu Glu Arg Gln Leu Asp Met Ser Leu Lys Gln Ile Arg
35 40 45

Ser Thr Arg Thr Gln Leu Met Leu Asp Gln Leu Gln Asp Leu Gln Arg
50 55 60

Lys
65

(25) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Gln Glu Tyr Leu Lys Leu Lys Gly Arg Tyr Glu Ala Leu Gln Arg Ser
5 10 15

Gln Arg Asn Leu Leu Gly Glu Asp Leu Gly Pro Leu Asn Ser Lys Glu
20 25 30

Leu Glu Ser Leu Glu Arg Gln Leu Asp Met Ser Leu Lys Gln Ile Arg
35 40 45

Ser Thr Arg Thr Gln Leu Met Leu Asp Gln Leu Thr Asp Tyr Gln Arg
50 55 60

Lys
65

(26) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Gln Glu Tyr Leu Lys Leu Lys Asn Arg Val Glu Ala Leu Gln Arg Ser
5 10 15

Gln Arg Asn Leu Leu Gly Glu Asp Leu Gly Pro Leu Gly Ser Lys Glu
20 25 30

Leu Glu Gln Leu Glu Arg Gln Leu Asp Ser Ser Leu Arg Gln Ile Arg
35 40 45

Ser Thr Arg Thr Gln Phe Met Leu Asp Gln Leu Ala Asp Leu Gln Arg
50 55 60

Arg
65

(27) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Arg Glu Tyr Leu Lys Leu Lys Gly Arg Tyr Glu Asn Leu Gln Arg Gln
5 10 15

Gln Arg Asn Leu Leu Gly Glu Asp Leu Gly Pro Leu Asn Ser Lys Glu
20 25 30

Leu Glu Gln Leu Glu Arg Gln Leu Asp Gly Ser Leu Lys Gln Val Arg
35 40 45

Ser Ile Lys Thr Gln Tyr Met Leu Asp Gln Leu Ser Asp Leu Gln Asn

50 55 60

Lys
65

(28) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Arg Glu Tyr Leu Lys Leu Lys Gly Arg Tyr Glu Asn Leu Gln Arg Gln
5 10 15

Gln Arg Asn Leu Leu Gly Glu Asp Leu Gly Pro Leu Asn Ser Lys Glu
20 25 30

Leu Glu Gln Leu Glu Arg Gln Leu Asp Gly Ser Leu Lys Gln Val Arg
35 40 45

Cys Ile Lys Thr Gln Tyr Met Leu Asp Gln Leu Ser Asp Leu Gln Gly
50 55 60

Lys
65

(29) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Met Glu Tyr Asn Arg Leu Lys Ala Lys Ile Glu Leu Leu Glu Arg Asn
5 10 15

Gln Arg His Tyr Leu Gly Glu Asp Leu Gln Ala Met Ser Pro Lys Glu
20 25 30

Leu Gln Asn Leu Glu Gln Gln Leu Asp Thr Ala Leu Lys His Ile Arg
35 40 45

Thr Arg Lys Asn Gln Leu Met Tyr Glu Ser Ile Asn Glu Leu Gln Lys
50 55 60

Lys
65

(30) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Gln Glu Ser Ala Lys Leu Arg Gln Gln Ile Ile Ser Ile Gln Asn Ser
5 10 15

Asn Arg Gln Leu Met Gly Glu Thr Ile Gly Ser Met Ser Pro Lys Glu
20 25 30

Leu Arg Asn Leu Glu Gly Arg Leu Glu Arg Ser Ile Thr Arg Ile Arg
35 40 45

Ser Lys Lys Asn Glu Leu Leu Phe Ser Glu Ile Asp Tyr Met Gln Lys
50 55 60

Arg
65

(31) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 66 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Gln Glu Thr Lys Arg Lys Leu Leu Glu Thr Asn Arg Asn Leu Arg Thr
5 10 15

Gln Ile Lys Gln Arg Leu Gly Glu Cys Leu Asp Glu Leu Asp Ile Gln
20 25 30

Glu Leu Arg Arg Leu Glu Asp Glu Met Glu Asn Thr Phe Lys Leu Val
35 40 45

Arg Glu Arg Lys Phe Lys Ser Leu Gly Asn Gln Ile Glu Thr Thr Lys
50 55 60

Lys Lys
65

(32) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 61 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Asn Glu Ile Asp Arg Ile Lys Lys Glu Asn Asp Ser Leu Gln Leu Glu
5 10 15

Leu Arg His Leu Lys Gly Glu Asp Ile Gln Ser Leu Asn Leu Lys Asn
20 25 30

Leu Met Ala Val Glu His Ala Ile Glu His Gly Leu Asp Lys Val Arg
35 40 45

Asp His Gln Met Glu Ile Leu Ile Ser Lys Arg Arg Asn
50 55 60

(33) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Glu Pro Thr Leu Gln Ile Gly Tyr Pro His His Gln Phe Pro Pro Pro
5 10 15

Glu Ala Val Asn Asn Ile Pro Arg Ser Ala Ala Thr Gly Glu Asn Asn
20 25 30

Phe Met Leu Gly Trp Val Leu
35

(34) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Glu Pro Thr Leu Gln Ile Gly Tyr Pro Pro His His Gln Phe Leu Pro
5 10 15

Ser Glu Ala Ala Asn Asn Ile Pro Arg Ser Pro Pro Gly Gly Glu Asn
20 25 30

Asn Phe Met Leu Gly Trp Val Leu
35 40

(35) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 38 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Glu Pro Phe Leu Gln Ile Gly Phe Gly Gln His Tyr Tyr Val Gly Gly
5 10 15

Glu Gly Ser Ser Val Ser Lys Ser Asn Val Ala Gly Glu Thr Asn Phe
20 25 30

Val Gln Gly Trp Val Leu
35

(36) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Glu Pro Thr Leu Gln Ile Gly Tyr Gln Asn Asp Pro Ile Thr Val Gly
5 10 15

Gly Ala Gly Pro Ser Val Asn Asn Tyr Met Ala Gly Trp Leu Pro
20 25 30

(37) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Glu Pro Thr Leu Gln Ile Gly Tyr Gln Asn Asp Pro Ile Thr Val Gly
5 10 15

Gly Ala Gly Pro Ser Val Asn Asn Tyr Met Ala Gly Trp Leu Pro
20 25 30

(38) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Glu Pro Thr Leu Gln Ile Gly Tyr His Ser Asp Ile Thr Met Ala Thr
5 10 15

Ala Thr Ala Ser Thr Val Asn Asn Tyr Met Pro Pro Gly Trp Leu Gly
20 25 30

(39) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Asn Pro Thr Leu Gln Met Gly Tyr Asp Asn Pro Val Cys Ser Glu Gln
5 10 15

Ile Thr Ala Thr Thr Gln Ala Gln Ala Gln Pro Gly Asn Gly Tyr Ile
20 25 30

Pro Gly Trp Met Leu
35

(40) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Asp Pro Thr Leu Gln Ile Gly Tyr Ser His Pro Val Cys Ser Glu Gln
5 10 15

Met Ala Val Thr Val Gln Gly Gln Ser Gln Gln Gly Asn Gly Tyr Ile
20 25 30

Pro Gly Trp Met Leu
35

(41) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Ser Pro Phe Leu Asn Met Gly Gly Leu Tyr Gln Glu Asp Asp Pro Met
5 10 15

Ala Met Arg Asn Asp Leu Glu Leu Thr Leu Glu Pro Val Tyr Asn Cys
20 25 30

Asn Leu Gly Cys Phe Ala Ala
35

(42) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Val Ala Ala Leu Gln Pro Asn Asn His His Tyr Ser Ser Ala Gly Arg
5 10 15

Gln Asp Gln Thr Ala Leu Gln Leu Val
20 25

(43) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Phe His Gln Asn His His His Tyr Tyr Pro Asn His Gly Leu His Ala
5 10 15

Pro Ser Ala Ser Asp Ile Ile Thr Phe His Leu Leu Glu
20 25

(44) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acid residues

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Val Ala Ala Leu Gln Pro Asn Leu Gln Glu Lys Ile Met Ser Leu Val
5 10 15

Ile Asp

WHAT IS CLAIMED IS:

1. An isolated nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ. ID NO:2, SEQ. ID NO:4, and SEQ. ID NO:6.

2. The nucleic acid of claim 1 that, when expressed in a transgenic plant containing said nucleic acid, causes the transgenic plant to exhibit at least one phenotype selected from the group consisting of: (i) diminished apical dominance, (ii) early flowering, (iii) altered daylength requirement for flowering; (iv) greater synchronization of flowering; and (v) relaxed vernalization requirement, compared to a nontransgenic control plant.

3. The nucleic acid of claim 2 wherein expression of the nucleic acid in the transgenic plant causes the transgenic plant to exhibit diminished apical dominance and early flowering compared to the nontransgenic control plant.

4. An isolated nucleic acid comprising at least 30 consecutive nucleotides of SEQ ID NO:2.

5. The nucleic acid of claim 4 comprising a portion of SEQ ID NO:2, wherein the portion being selected from the group consisting of a protein-coding region, a MADS-box region, and a K-box region.

6. An isolated nucleic acid comprising at least 100 consecutive nucleotides having at least 70% nucleotide sequence similarity with a nucleotide sequence of SEQ ID NO:2, not including MADS-box and K-box regions thereof.

7. The nucleic acid of claim 6, wherein expression of the nucleic acid in a transgenic plant causes the transgenic plant to exhibit at least one phenotype in the transgenic plant selected from the group consisting of: (i) diminished apical dominance, (ii) early flowering, (iii) altered daylength requirement for flowering; (iv) greater synchronization of flowering; and (v) relaxed vernalization requirement compared to a nontransgenic control plant.

8. The nucleic acid of claim 7 wherein expression of the nucleic acid in the transgenic plant causes the transgenic plant to exhibit diminished apical dominance and early flowering compared to the nontransgenic control plant.

9. The nucleic acid of claim 7 comprising at least 100 consecutive nucleotides that comprises only silent or conservative substitutions to the nucleotide sequence of SEQ ID NO:2.

10. A transgenic plant comprising the nucleic acid of claim 1.

11. A transgenic plant comprising the nucleic acid of claim 4.

12. A transgenic plant comprising the nucleic acid of claim 6.

13. A method for producing a transgenic plant, comprising the steps of:

(a) providing a nucleic acid as recited in claim 1;

(b) introducing the nucleic acid of step (a) into a plant cell, thereby producing a transformed plant cell; and

(c) regenerating from the transformed plant cell a transgenic plant comprising the nucleic acid.

14. A method for producing a transgenic plant, comprising the steps of:

(a) providing a nucleic acid as recited in claim 4;

(b) introducing the nucleic acid of step (a) into a plant cell, thereby producing a transformed plant cell; and

(c) regenerating from the transformed plant cell a transgenic plant comprising the nucleic acid.

15. A method for producing a transgenic plant, comprising the steps of:

(a) providing a nucleic acid as recited in claim 6;

(b) introducing the nucleic acid of step (a) into a plant cell, thereby producing a transformed plant cell; and

(c) regenerating from the transformed plant cell a transgenic plant comprising the nucleic acid.

FIG. 1

```
   1 TACCCGCGGGAATCGTTCGATCGATCGGGCGAGATGGGGAGGGGAAGAGTTGAGCTGAAG
                                       M  G  R  G  R  V  E  L  K      9

  61 CGCATCGAGAACAAGATCAACAGGCAGGTCACCTTCTCCAAGCGCCGCAACGGCCTCCTC
     R  I  E  N  K  I  N  R  Q  V  T  F  S  K  R  R  N  G  L  L     29
 121 AAGAAGGCCTACGAGCTGTCCGTTCTCTGCGACGCCGAGGTCGCGCTCATCATCTTCTCC  MADS-box
     K  K  A  Y  E  L  S  V  L  C  D  A  E  V  A  L  I  I  F  S     49
 181 AGCCGCGGCAAGCTCTACGAGTTCGGCAGCGCCGGCATAACAAAGACTTTAGAAAGGTAC
     K  S  R  G  L  Y  E  F  G  S  A  G  I  T  K  T  L  E  R  Y     69
 241 CAACATTGTTGCTACAATGCTCAAGATTCCAACAATGCACTTTCTGAAACTCAGAGTTGG
     Q  H  C  C  Y  N  A  Q  D  S  N  N  A  L  S  E  T  Q  S  W     89
 301 TACCATGAAATGTCAAAGTTGAAAGCAAAATTTGAAGCTTTGCAGCGCACTCAAAGGCAC
     Y  H  E  M  S  K  L  K  A  K  F  E  A  L  Q  R  T  Q  R  H    109
 361 TTGCTTGGGGAGGATCTTGGACCACTCAGCGTCAAAGAATTGCAGCAGCTGGAGAAACAG  K-box
     L  L  G  E  D  L  G  P  L  S  V  K  E  L  Q  Q  L  E  K  Q    129
 421 CTTGAATGTGCACTATCACAGGCGAGACAGAGAAAGACGCAACTGATGATGGAACAGGTG
     L  E  C  A  L  S  Q  A  R  Q  R  K  T  Q  L  M  M  E  Q  V    149
 481 GAGGAACTTCGCAGAAAGGAGCGTCAGCTGGGTGAAATTAATAGGCAACTCAAGCACAAG
     E  E  L  R  R  K  E  R  Q  L  G  E  I  N  R  Q  L  K  H  K    169
 541 CTCGAGGTTGAAGGTTCCACCAGCAACTACAGAGCCATGCAGCAAGCCTCCTGGGCTCAG
     L  E  V  E  G  S  T  S  N  Y  R  A  M  Q  Q  A  S  W  A  Q    189
 601 GGCGCCGTGGTGGAGAATGGCGCCGCATACGTGCAGCCGCCGCCACACTCCGCGGCCATG
     G  A  V  V  E  N  G  A  A  Y  V  Q  P  P  P  H  S  A  A  M    209
 661 GACTCTGAACCCACCTTGCAAATTGGGTATCCTCATCAATTTGTGCCTGCTGAAGCAAAC
     D  S  E  P  T  L  Q  I  G  Y  P  H  Q  F  V  P  A  E  A  N    229
 721 ACTATTCAGAGGAGCACTGCCCCTGCAGGTGCAGAGAACAACTTCATGCTGGGATGGGTT
     T  I  Q  R  S  T  A  P  A  G  A  E  N  N  F  M  L  G  W  V    249
 781 CTTTGAGCTAAGCAGCCATCGATCAGCTGTCAGAAGTTGGAGCTAATAATAAAAGGGATG
     L  *                                                           250
 841 TGGAGTGGGCTACATGTATCTCGGATCTCTCTGCGAGCCACCTAATGGTCTTGCGTGGCC
 901 CTTTAATCTGTATGTTTTTGTGTGTAAGCTACTGCTAGCTGTTTGCACCTTCTGCGTCCG
 961 TGGTTGTGTTTCCGTGCTACCTTTTTATGTTTTGATTTGGATCTTGTTTGAAAATAATCT
1021 TACCAGCTTTGGGTAAACTGTTT(A)n
```

FIG. 2

```
   1  TATCCCCTTCCTCCAGGTGGCTTGTTTCTTGCAGTGGTGGTGGTGGTGGTGGTGAGATCT
  61  AGCTTGGTTGGTTGGTGGCAGCTGGAGATCGATCGGGATGGGGAGGGGGCGGGTGGAGCT
                                          M  G  R  G  R  V  E  L    8
                                                                  MADS-box
 121  GAAGAGGATCGAGAACAAGATCAACCGGAAGGTGACGTTCGCCAAGAGGAGGAATGGCCT
       K  R  I  E  N  K  I  N  R  K  V  T  F  A  K  R  R  N  G  L   28
 181  GCTCAAGAAGGCGTACGAGCTCTCCGTCCTCTGCGACGCCGAGGTCGCCCTCATCATCTT
       L  K  K  A  Y  E  L  S  V  L  C  D  A  E  V  A  L  I  I  F   48
 241  CTCCAACCGCGGCAAGCTCTACGAGTTCTGCAGCACCCAGAGCATGACTAAAACGCTTGA
       S  N  R  G  K  L  Y  E  F  C  S  T  Q  S  M  T  K  T  L  E   68
 301  GAAGTATCAGAAATGCAGTTACGCAGGACCCGAAACAGCTGTCCAAAATAGAGAAAGTGA
       K  Y  Q  K  C  S  Y  A  G  P  E  T  A  V  Q  N  R  E  S  E   88
 361  GCAATTGAAAGCTAGCCGCAATGAATACCTCAAACTGAAGGCAAGGGTTGAAAATTTACA
       Q  L  K  A  S  R  N  E  Y  L  K  L  K  A  R  V  E  N  L  Q  108
                                                                  K BOX
 421  ACGGACTCAAAGAAATTTGCTGGGTCCAGATCTTGATTCATTAGGCATAAAAGAGCTCGA
       R  T  Q  R  N  L  L  G  P  D  L  D  S  L  G  I  K  E  L  E  128
 481  GAGCCTAGAGAAGCAGCTTGATTCATCCCTGAAGCACGTCAGAACTACAAGGACAAAACA
       S  L  E  K  Q  L  D  S  S  L  K  H  V  R  T  T  R  T  K  H  148
 541  TCTGGTCGACCAACTGACGGAGCTTCAGAGAAAGGAACAAATGGTTTCTGAAGCAAATAG
       L  V  D  Q  L  T  E  L  Q  R  K  E  Q  M  V  S  E  A  N  R  168
 601  ATGCCTTAGGAGAAAACTGGAGGAAAGCAACCATGTTCGCGGGCAGCAAGTGTGGGAGCA
       C  L  R  R  K  L  E  E  S  N  H  V  R  G  Q  Q  V  W  E  Q  188
 661  GGGCTGCAACTTAATTGGCTATGAACGTCAGCCTGAAGTGCAGCAGCCTCTTCACGGCGG
       G  C  N  L  I  G  Y  E  R  Q  P  E  V  Q  Q  P  L  H  G  G  208
 721  CAATGGGTTCTTCCATCCACTTGATGCTGCTGGTGAACCCACCCTTCAGATTGGGTACCC
       N  G  F  F  H  P  L  D  A  A  G  E  P  T  L  Q  I  G  Y  P  228
 781  TGCAGAGCATCATGAGGCGATGAACAGTGCGTGCATGAACACCTACATGCCCCCATGGCT
       A  E  H  H  E  A  M  N  S  A  C  M  N  T  Y  M  P  P  W  L  248
 841  ACCATGATGATGACGGGACAATGAATTACGAAATAACAAGGATATGTGGCATATATGTGG
       P  *                                                        249
 901  TGCCGCATACATGCATGTATCATGGCTAGCTACTTAATTGGAGTGATGGATTTGAACTAG
 961  TTTCGTATGTAGCCTGTTTGTGTGTAACTTGTGTGAGATACTACCTTAAAAACTATCGGT
1021  GTCTGTTGAACATATTCTGCGATCAACTTTAAGCGTATT(A)n
```

FIG. 3

```
   1  TGCTTTCCCCTCTCTTCCGCTTCGCGAGATTGGTTGATTCATCTCGCGATTGATCGAGCT
  61  CGAGCGGCGGTGAGGTGAGGTGGAGGAGGAGGAGGAGGAGGAGATCGGGATGGGGAGAGG
                                                    M  G  R  G      4
 121  GAGGGTGGAGCTGAAGAGGATCGAGAACAAGATCAACAGGCAGGTGACGTTCGCGAAGCG   MADS-box
       R  V  E  L  K  R  I  E  N  K  I  N  R  Q  V  T  F  A  K  R     24
 181  GAGGAATGGGCTGCTCAAGAAGGCGTACGAGCTCTCCGTGCTCTGCGACGCCGAGGTCGC
       R  N  G  L  L  K  K  A  Y  E  L  S  V  L  C  D  A  E  V  A     44
 241  CCTCATCATCTTCTCCAACCGCGGCAAGCTCTACGAGTTCTGCAGCGGCCAAAGCATGAC
       L  I  I  F  S  N  R  G  K  L  Y  E  F  C  S  G  Q  S  M  T     64
 301  CAGAACTTTGGAAAGATACCAAAAATTCAGTTATGGTGGGCCAGATACTGCAATACAGAA
       R  T  L  E  R  Y  Q  K  F  S  Y  G  G  P  D  T  A  I  Q  N     84
 361  CAAGGAAAATGAGTTAGTGCAAAGCAGCCGCAATGAGTACCTCAAACTGAAGGCACGGGT
       K  E  N  E  L  V  Q  S  S  R  N  E  Y  L  K  L  K  A  R  V    104
 421  GGAAAATTTACAGAGGACCCAAAGGAATCTTCTTGGTGAAGATCTTGGGACACTTGGCAT   K-box
       E  N  L  Q  R  T  Q  R  N  L  L  G  E  D  L  G  T  L  G  I    124
 461  AAAAGAGCTAGAGCAGCTTGAGAAACAACTTGATTCATCCTTGAGGCACATTAGATCCAC
       K  E  L  E  Q  L  E  K  Q  L  D  S  S  L  R  H  I  R  S  T    144
 541  AAGGACACAGCATATGCTTGATCAGCTCACTGATCTCCAGAGGAGGGAACAAATGTTGTG
       R  T  Q  H  M  L  D  Q  L  T  D  L  Q  R  R  E  Q  M  L  C    164
 601  TGAAGCAAATAAGTGCCTCAGAAGAAAACTGGAGGAGAGCAACCAGTTGCATGGACAAGT
       E  A  N  K  C  L  R  R  K  L  E  E  S  N  Q  L  H  G  Q  V    184
 661  GTGGGAGCACGGCGCCACCCTACTCGGCTACGAGCGGCAGTCGCCTCATGCCGTCCAGCA
       W  E  H  G  A  T  L  L  G  Y  E  R  Q  S  P  H  A  V  Q  Q    204
 721  GGTGCCACCGCACGGTGGCAACGGATTCTTCCATTCCCTGGAAGCTGCCGCCGAGCCCAC
       V  P  P  H  G  G  N  G  F  F  H  S  L  E  A  A  A  E  P  T    224
 781  CTTGCAGATCGGGTTTACTCCAGAGCAGATGAACAACTCATGCGTGACTGCCTTCATGCC
       L  Q  I  G  F  T  P  E  Q  M  N  N  S  C  V  T  A  F  M  P    244
 841  GACATGGCTACCCTGAACTCCTGAAGGCCGATGCGACAACCAATAAAAACGGATGTGACG
       T  W  L  P  *                                                  248
 901  ACACAGATCAAGTCGCACCATTAGATTGATCTTCTCCTACAAGAGTGAGACTAGTAATTC
 961  CGCGTTTGTGTGCTAGCGTGTTGAAACTTTTCTGATGTGATGCACGCACTTTTAATTATT
1021  ATTAAGCGTTCAAGGACTAGTATGTGGTATAAAAGCCCGTACGTGACAGCCTATGGTTAT
1081  ATGCTGCGCAAAAACTACGTATGGTACAGTGCAGTGCCTGTACATTTCATAATTTGCGGG
1141  TAAAGTTTATTGACTATATATCCAGTGTGTCAAATATAAT(A)n
```

FIG. 4A

MADS BOX

GRGRVELKRIENKINRQVTFSKRRNGLLKKAYELSVLCDAEVALIIFSSRGKLYEF	1 - OsMADS6
GRGRVELKRIENKINRQVTFSKRRNGLLKKAYELSVLCDAEVALIIFSSRGKLYEF	2 - ZAG3
GRGRVELKRIENKINRQVTFSKRRNGLLKKAYELSVLCDAEVALIIFSGRGKLYEF	3 - ZAG5
GRGRVEMKRIENKINRQVTFSKRRNGLLKKAYELSVLCDAEVALIIFSSRGKLYEF	4 - AGL6
GRGRVELKRIENKINRQVTFAKRRNGLLKKAYELSVLCDAEVALIIFSNRGKLYEF	5 - OsMADS8
GRGRVELKRIENKINRQVTFAKRRNGLLKKAYELSVLCDAEVALIIFSNRGKLYEF	6 - OsMADS7
GRGRVELKRIENKINRQVTFAKRRNGLLKKAYELSVLCDAEVALIIFSNRGKLYEF	7 - FBP2
GRGRVELKRIEGKINRQVTFAKRRNGLLKKAYELSVLCDAEVALIIFSNRGKLYEF	8 - TM5
GRGRVELKMIENKINRQVTFAKRRKRLLKKAYELSVLCDAEVALIIFSNRGKLYEF	9 - OM1
GRGRVELKRIENKINRQVTFAKRRNGLLKKAYELSVLCDAEVALIIFSNRGKLYEF	10 - AGL2
GRGRVELKRIENKINRQVTFAKRRNGLLKKAYELSVLCDAEVSLIVFSNRGKLYEF	11 - AGL4
GRGKVELKRIENKISRQVTFAKRRNGLLKKAYELSLLCDAEVALIIFSGRGRLFEF	12 - OsMADS1
GRGRVQLKRIENKINRQVTFSKRRAGLLKKAHEISVLCDAEVALVVFSHKGKLFEY	13 - AP1
GRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVALIVFSSRGRLYEY	14 - AG
ARGKIQIKRIENQTNRQVTYSKRRNGLFKKAHELTVLCDARVSIIMFSSSNKLHEY	15 - AP3
GRGKIEIKRIENANNRVVTFSKRRNGLVKKAKEITVLCDAKVALIIFASNGKMIDY	16 - PI

FIG. 4B

K BOX

HEMSKLKAKFEALQRTQ.RHLLGEDLGPLSVKELQQLEKQLECALSQARQRKTQLMMEQVEELRRK	1
QEMSKLRAKFEALQRTQ.RHLLGEELGPLSVKELQQLEKQLECALSQARQRKTQLMMEQVEELRRK	2
QEMSKLRAKFEALQRTQ.RHLLGEDLGPLSVKELQQLEKQLECALSQARQRKTQVMMEQVEELRRT	3
QEVTKLKSKYESLVRTN.RNLLGEDLGEMGVKELQALERQLEAALTATRQRKTQVMMEEMEDLRKK	4
NEYLKLKARVENLQRTQ.RNLLGEDLGTLGIKELEQLEKQLDSSLRHIRSTRTQHMLDQLTDLQRR	5
NEYLKLKARVENLQRTQ.RNLLGPDLDSLGIKELESLEKQLDSSLKHVRTTRTKHLVDQLTELQRK	6
QEYLKLKARYEALQRSQ.RNLLGEDLGPLNSKELESLERQLDMSLKQIRSTRTQLMLDQLQDLQRK	7
QEYLKLKGRYEALQRSQ.RNLLGEDLGPLNSKELESLERQLDMSLKQIRSTRTQLMLDQLTDYQRK	8
QEYLKLKNRVEALQRSQ.RNLLGEDLGPLGSKELEQLERQLDSSLRQIRSTRTQFMLDQLADLQRR	9
REYLKLKGRYENLQRQQ.RNLLGEDLGPLNSKELEQLERQLDGSLKQVRSIKTQYMLDQLSDLQNK	10
REYLKLKGRYENLQRQQ.RNLLGEDLGPLNSKELEQLERQLDGSLKQVRCIKTQYMLDQLSDLQGK	11
QEYLKLKTRVEFLQTTQ.RNILGEDLGPLSMKELEQLENQIEVSLKQIRSRKNQALLDQLFDLKSK	12
MEYNRLKAKIELLER.NQRHYLGEDLQAMSPKELQNLEQQLDTALKHIRTRKNQLMYESINELQKK	13
QESAKLRQQI.ISIQNSNRQLMGETIGSMSPKELRNLEGRLERSITRIRSKKNELLFSEIDYMQKR	14
QETKRKLLETNRNLRTQIKQRLGECLDELDIQELRRLEDEMENTFKLVRERKFKSLGNQIETTKKK	15
NE.IDRIKKENDSLQLELRHLKGE...DIQSLNLKNLMAVEHAIEHGLDKVRDHQMEILISKRRN.	16

FIG. 4C C-TERMINAL END

EPTLQIGYP.H.QFVPAEANTIQRSTAPAGAENNFML.GWVL	1
EPTLQIGYP.HHQFPPPEA.VNNIPRSAATGENNFML.GWVL	2
EPTLQIGYPPHHQFLPSEA.ANNIPRSPPGGENNFML.GWVL	3
EPFLQIG...FGQHYYVGGEGSSVSKSNVAGETNFVQ.GWVL	4
EPTLQIG.....FTPEQMNNS.CVTAFMPT.WLP	5
EPTLQIGY.....PAEHHEAMNSACMNTYMPP.WLP	6
EPTLQIGY.....QNDPITVGGAGPSVNNYMAG.WLP	7
EPTLQIGY.....QNDPITVGGAGPSVNNYMAG.WLP	8
EPTLQIGY.....HSDITMATATASTVNNYMPPGWLG+7	9
NPTLQMGY....DNPVCSEQITATTQAQAQPGNGYIP.GWML	10
DPTLQIGY....SHPVCSEQMAVTVQGQSQQGNGYIP.GWML	11
DHSLQIGYHHPHAHHHQAYMDHLSNEAADMVAHHPNEHIPSG+2	12
SPFLNMGGLYQEDDPMAMRNDLELTLEPVYNCNLGCFAA	13
VAALQPNNHHYSSAGRQDQTALQLV	14
FHQNHHHYYPNHGLHAPSASDIITFHLLE	15
VAALQPNLQEKIMSLVID	16

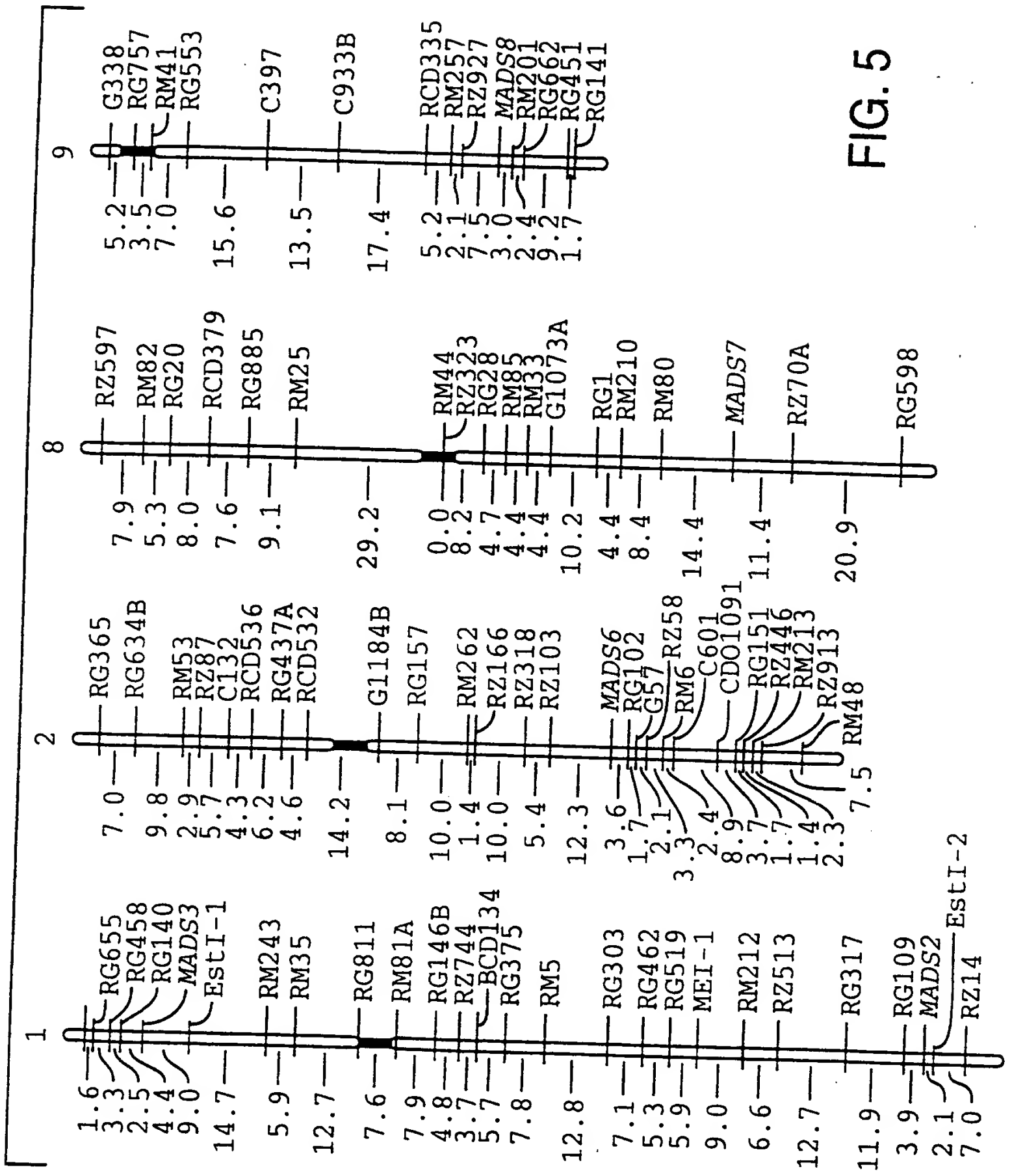

9
G338 5.2
RG757 3.5
RM41 7.0
RG553 15.6
C397 13.5
C933B 17.4
RCD335 5.2
RM257 2.1
RZ927 7.5
MADS8 3.0
RM201 2.4
RG662 9.2
RG451 1.7
RG141
8
RZ597 7.9
RM82 5.3
RG20 8.0
RCD379 7.6
RG885 9.1
RM25 29.2
RM44 0.0
RZ323 8.2
RG28 4.7
RM85 4.4
RM33 4.4
G1073A 10.2
RG1 4.4
RM210 8.4
RM80 14.4
MADS7 11.4
RZ70A 20.9
RG598
2
RG365 7.0
RG634B 9.8
RM53 2.9
RZ87 5.7
C132 4.3
RCD536 6.2
RG437A 4.6
RCD532 14.2
G1184B 8.1
RG157 10.0
RM262 1.4
RZ166 10.0
RZ318 5.4
RZ103 12.3
MADS6 3.6
RG102 1.7
G57 2.1
RZ58 3.3
RM6 2.4
C601 8.9
CDO1091 3.7
RG151 1.7
RZ446 1.4
RM213 2.3
RZ913 7.5
RM48
1
RG655 1.6
RG458 3.3
RG140 2.5
MADS3 4.4
EstI-1 9.0
RM243 14.7
RM35 5.9
RG811 12.7
RM81A 7.6
RG146B 7.9
RZ744 4.8
BCD134 3.7
RG375 5.7
RM5 7.8
RG303 12.8
RG462 7.1
RG519 5.3
MEI-1 5.9
RM212 9.0
RZ513 6.6
RG317 12.7
RG109 11.9
MADS2 3.9
EstI-2 2.1
RZ14 7.0

FIG. 5